NGI Program at DARPA

PITAC Review

Jan 14, 2000

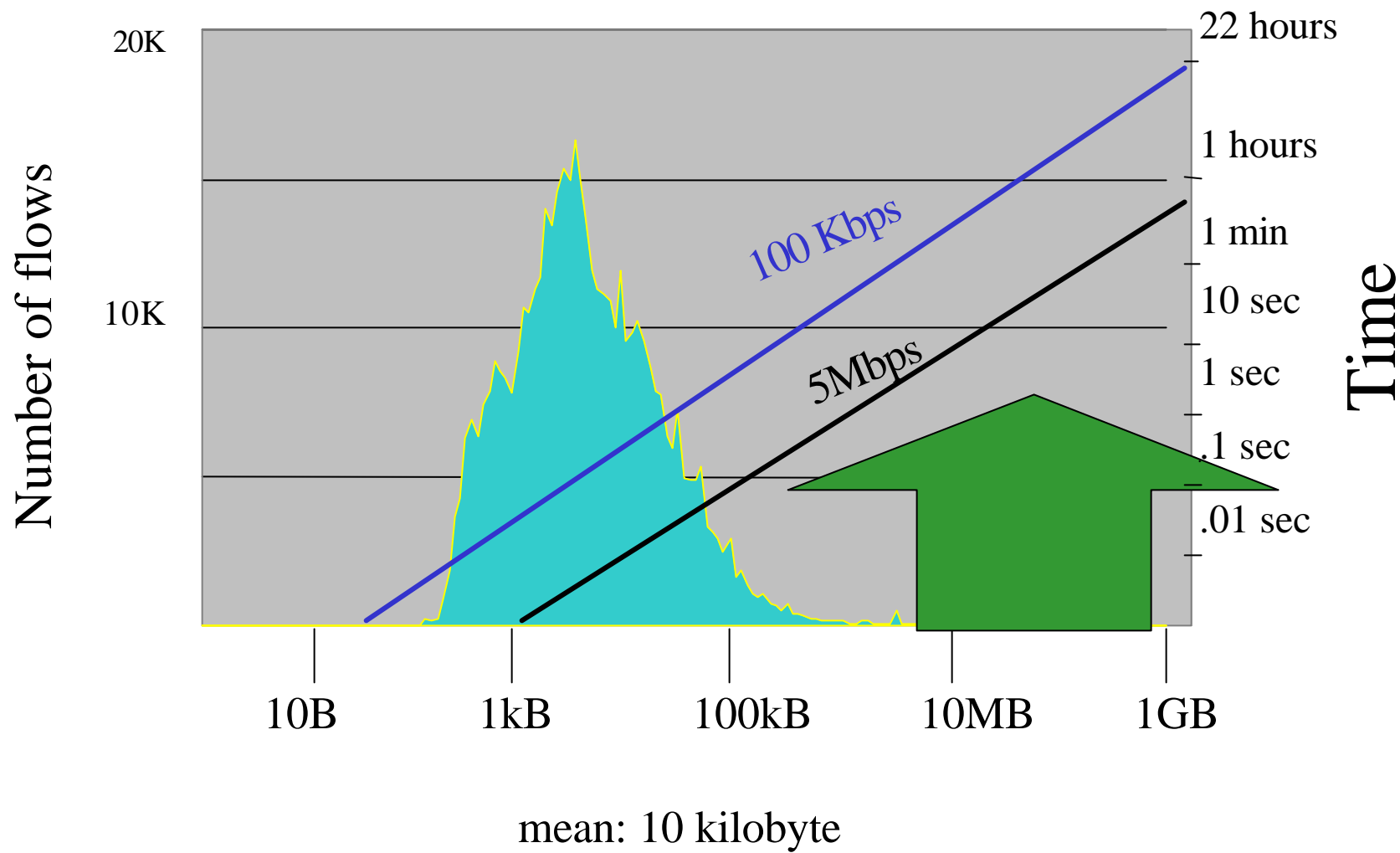Mari Maeda

DARPA
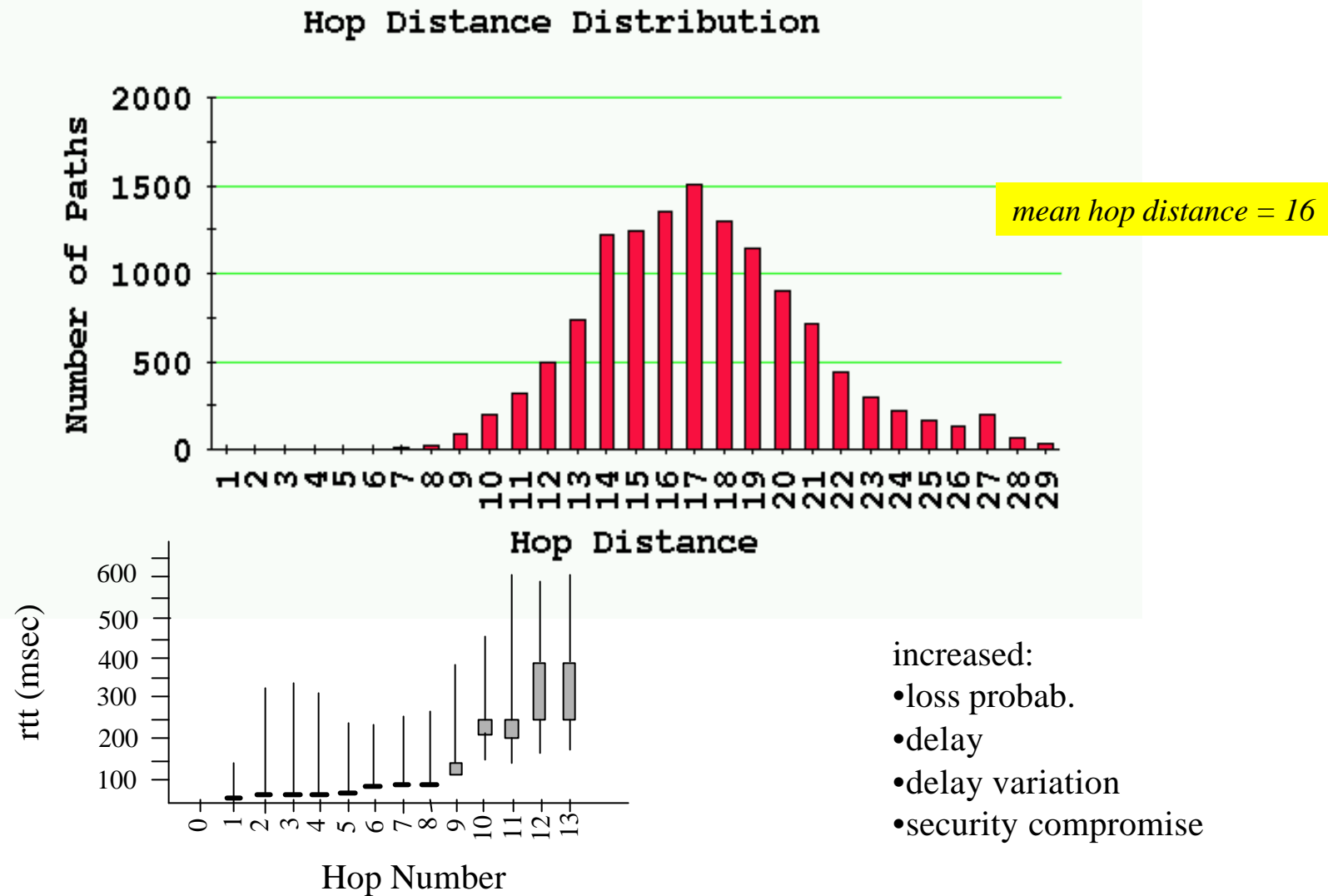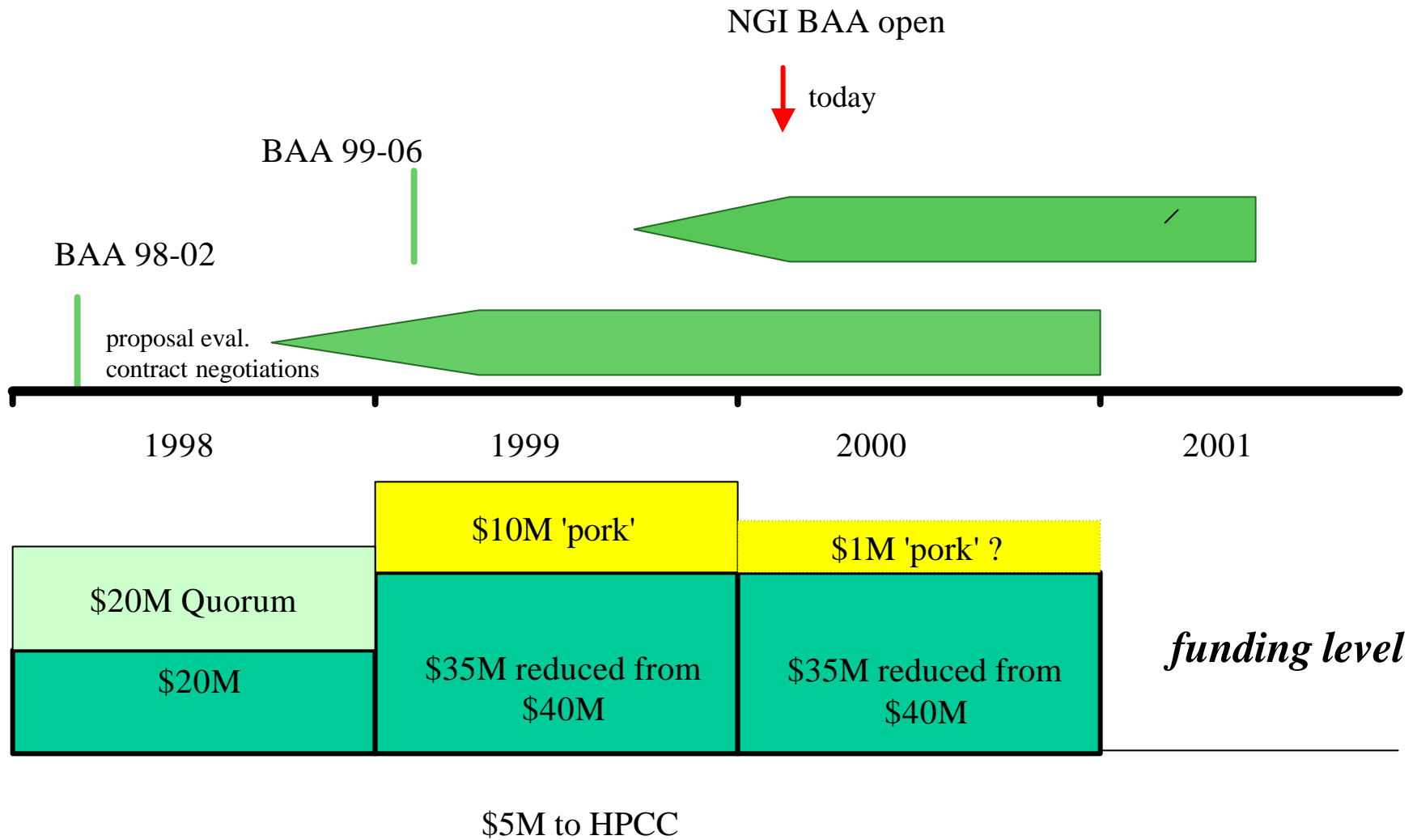
mmaeda@darpa.mil

Outline

- Introduction and Goals
- NGI Research
- SuperNet Testbed Status
- New Applications

Scaling the Internet

Hop Distance Distribution

mean hop distance = 16

increased:
- loss probab.
- delay
- delay variation
- security compromise

Program Time Line

NGI BAA open

today

BAA 99-06

BAA 98-02

proposal eval.
contract negotiations

| 1998 | 1999 | 2000 | 2001 |

$10M 'pork'

$1M 'pork' ?

$20M Quorum

$20M

$35M reduced from
$40M

$35M reduced from
$40M

funding level

$5M to HPCC

NGI Program Components

To enable <u>ultra-high bandwidth on demand</u> over national networks guaranteed over the shared infrastructure

> •*Simplified protocol layering - IP over dynamic Optical Network.*
>
> •*End-to-end performance*
>
> •*Testbed*

SuperNet Technology

Create tools that greatly <u>automate</u> planning and management functions enabling networks to grow while limiting the cost and complexity of network management and control

> •*Adaptive network management and control software*
>
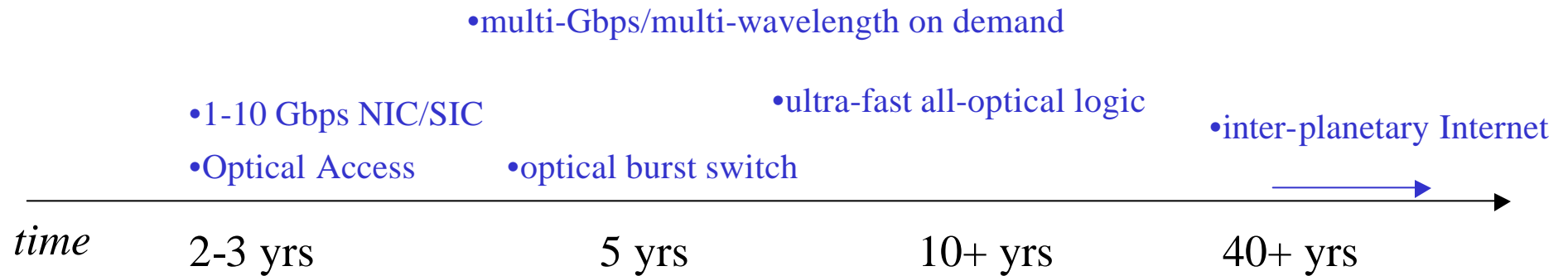> •*Large-scale network monitoring/analysis/visualization tools*

Network Engineering

Develop, test, deploy applications requiring gigabit end to end throughput
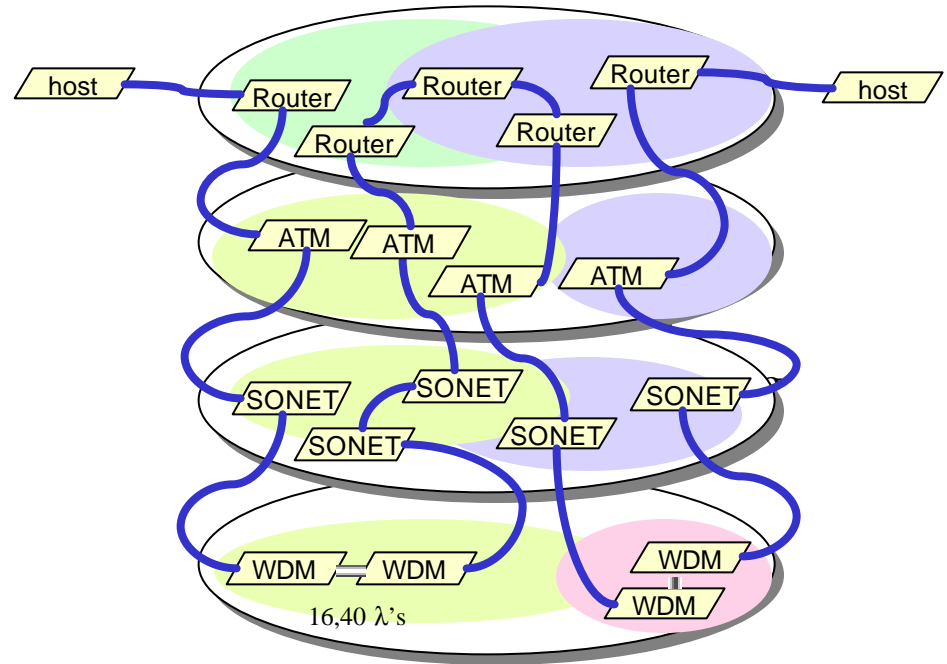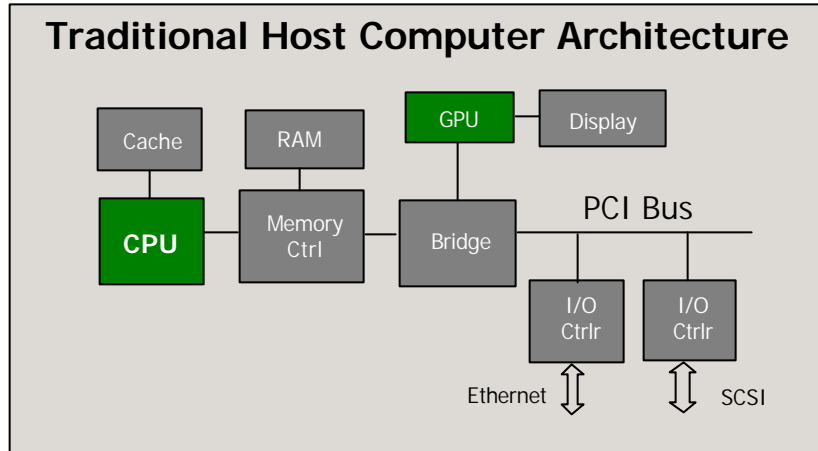
Applications

Diversity and number of end devices

Deeply Networked Systems

A range of projects addressing a range of near-term and long-term problems

•multi-Gbps/multi-wavelength on demand

•1-10 Gbps NIC/SIC

•ultra-fast all-optical logic

•inter-planetary Internet

•Optical Access   •optical burst switch

time

2-3 yrs   5 yrs   10+ yrs   40+ yrs

Environment is very dynamic; industry is moving very rapidly.
Goals and deliverables need to be adjusted accordingly.
(Surprise us.)

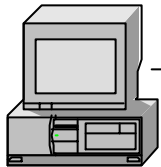# Gbps++ End to End Delivery

Traditional Host Computer Architecture

Cache, RAM, GPU, Display, CPU, Memory Ctrl, Bridge, PCI Bus, I/O Ctrlr, I/O Ctrlr, Ethernet, SCSI

host — Router — Router — Router — host
Router — Router

ATM — ATM — ATM — ATM

SONET — SONET — SONET — SONET

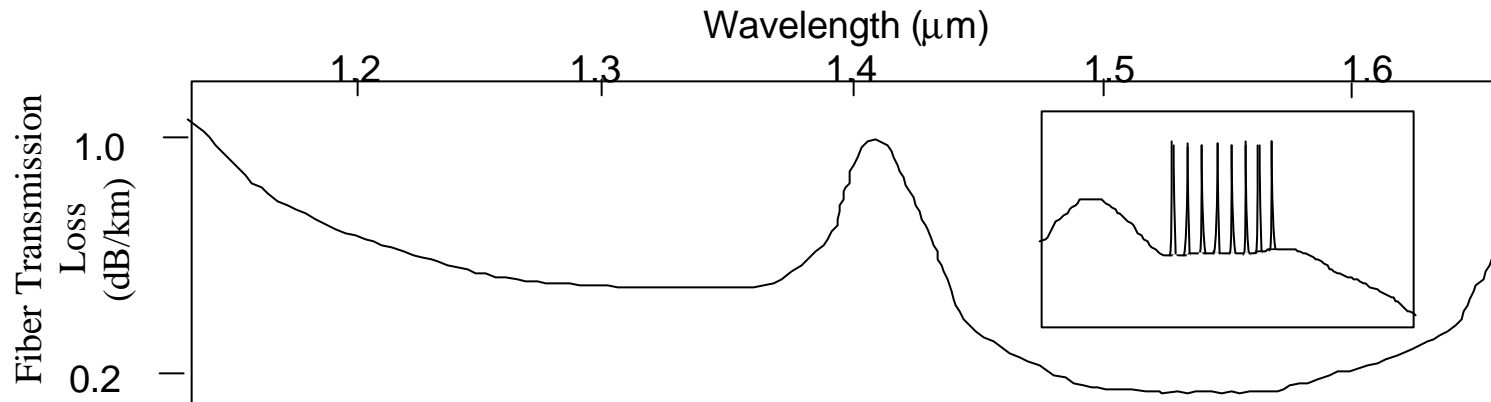WDM — WDM — WDM — WDM

16,40 λ's

**LAN**   **metro/regional**   **backbone**

- end-host architecture
- local/metro/regional network architecture (ONRAMP, HELIOS)
- protocol research (gbps tcpip, flow switch to bypass routers..)
- optical burst switch, label switch, packet switch
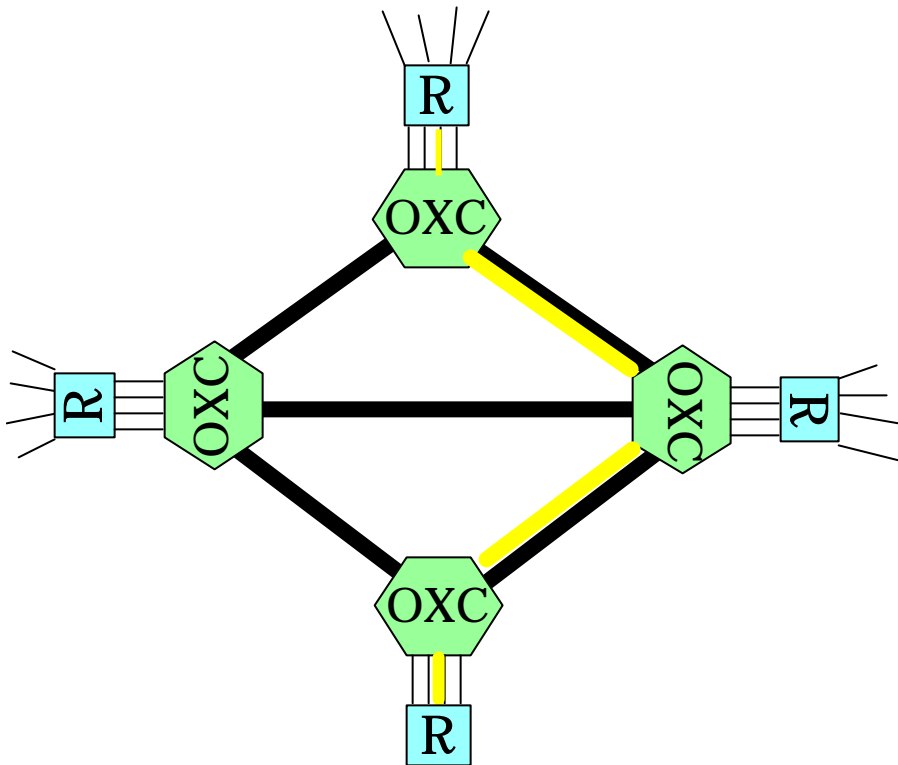
Wavelength Division Multiplexing

•**Savings in regenerator equipment and new fiber build costs**
•**Stretch capacity per fiber (#, rate per channel) and unregenerated transmission span**
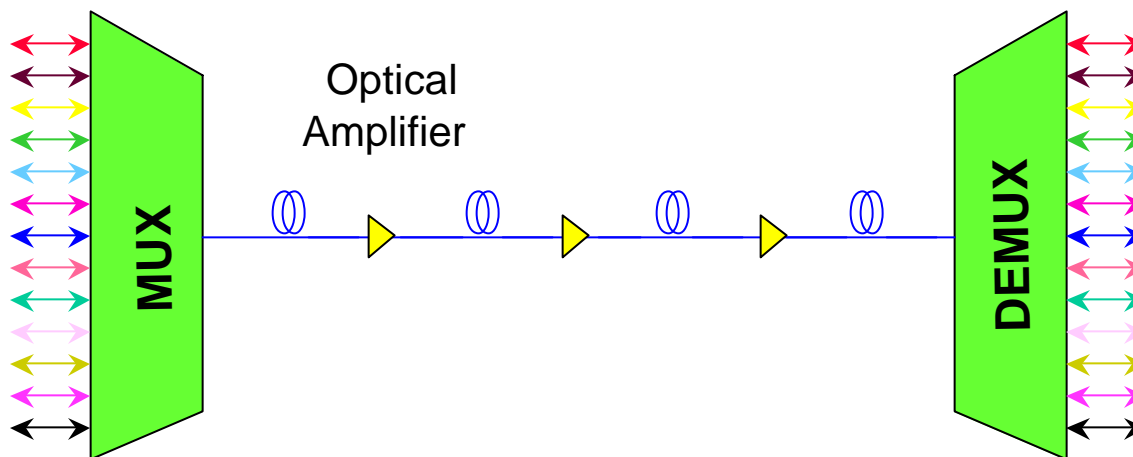
•fiber
•gain-flattened amplifier
•sources
•modulation techniques
•multiplexing techniques
•nonlinearities
•dispersion compensation

Optical Networking

• Dynamically configure new lightpaths to optically switch long sessions

Router initiated optical flow

courtesy of Adel Saleh
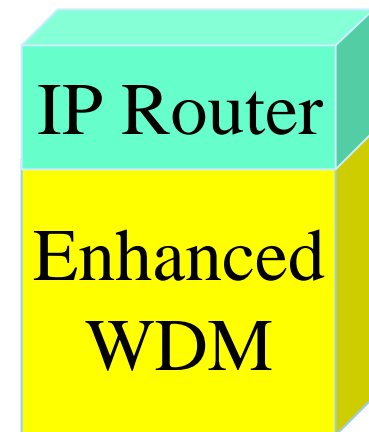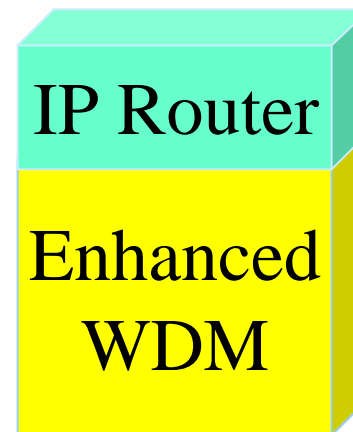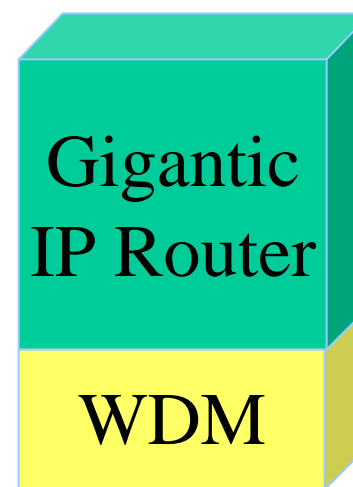
BIT Program'93-'99: Optical Networking

becoming commercialized, largely by startups
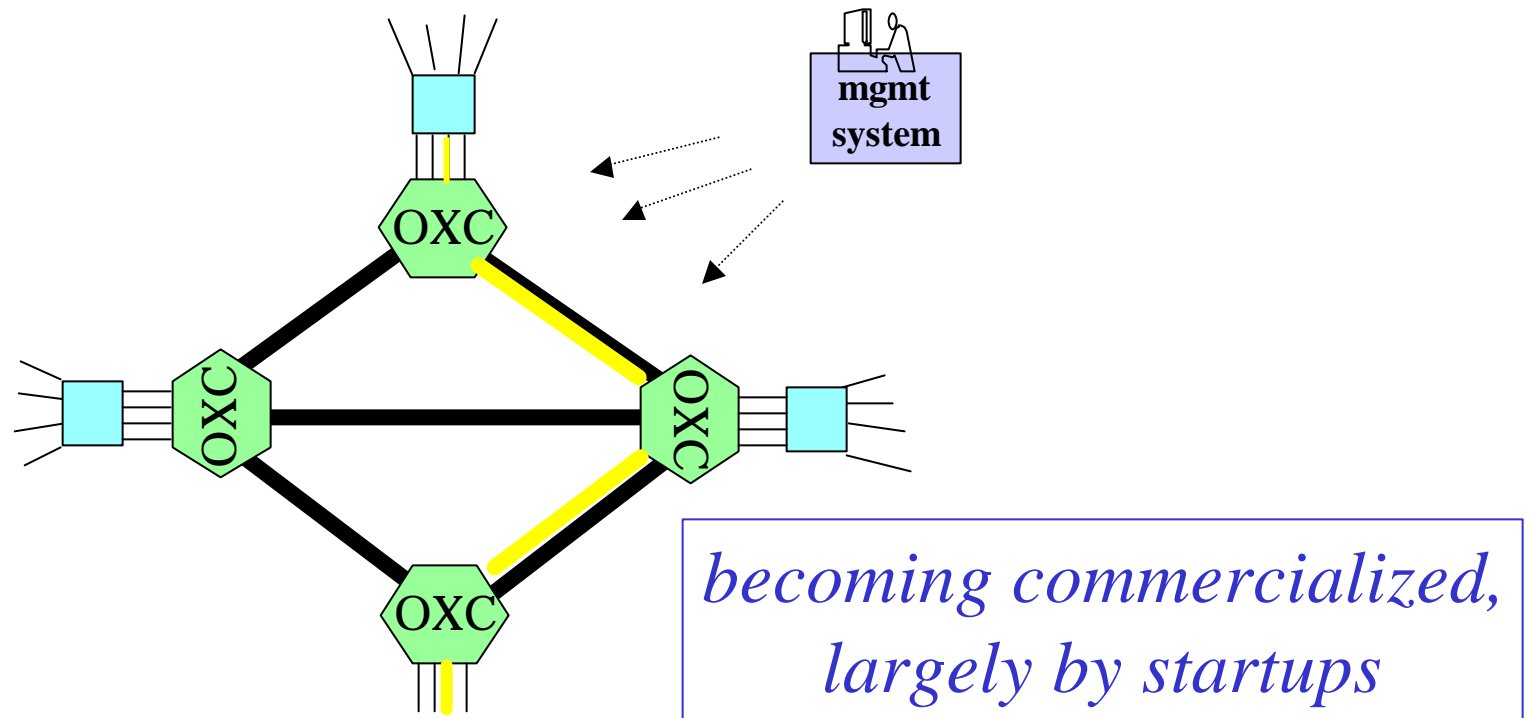
- Optical path set up by NC&M system
- Optical layer restoration
  (path vs line switched rings; mesh restoration)

Optical Flow Switching

today's
IP / WDM

router to router
flow switching

end to end
flow switching
(1-10 Gbps end to end)
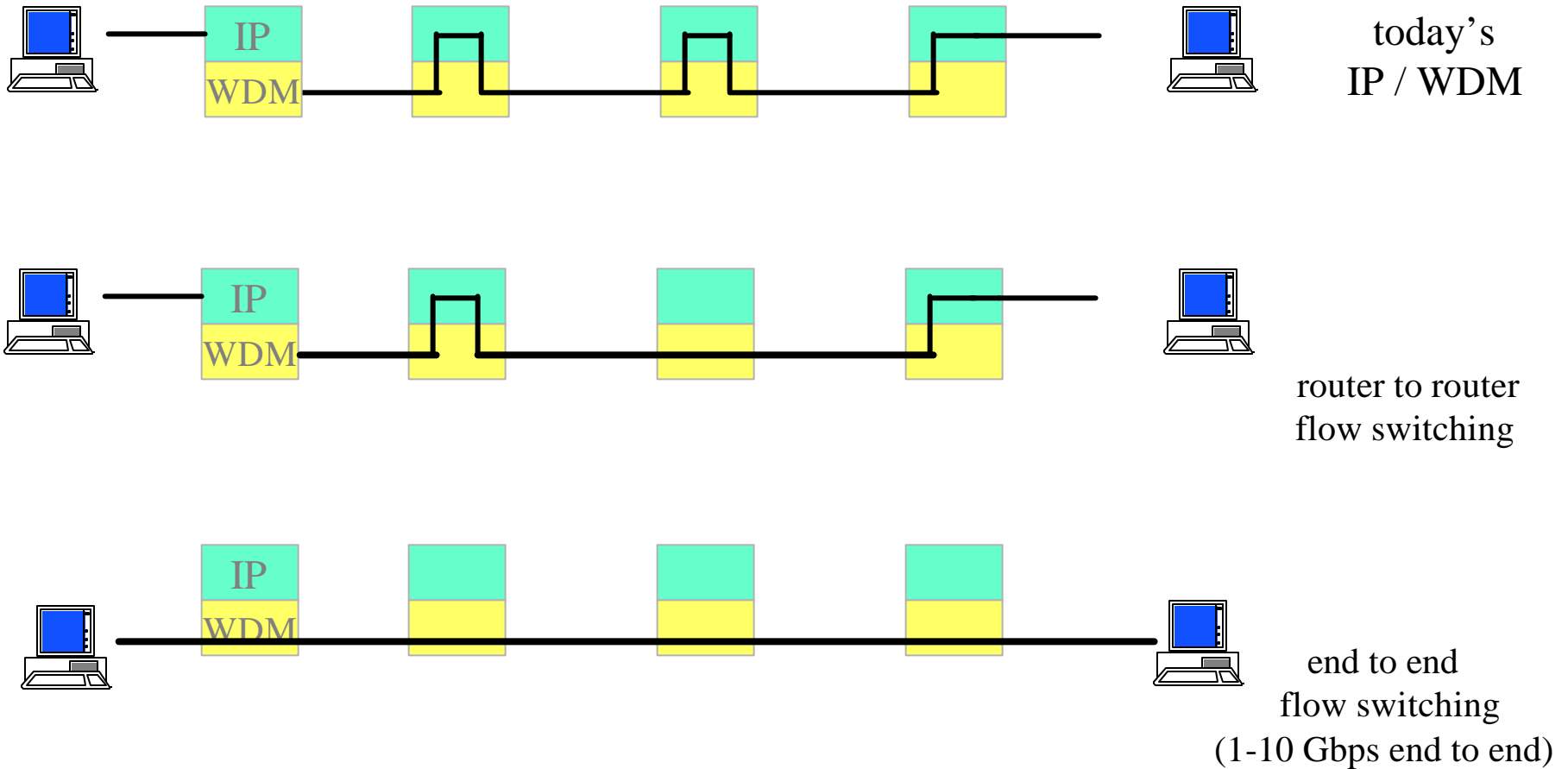
- End-Host triggered Optical Flow Switching
(1 - 10 Gbps end-to-end)

Advanced Optical Networking
(bypassing/offloading electronics)

holding time

Reconfigurable Optical Networking >minutes

Optical Flow Switching   >100 msec
(router or end-host triggered)

Optical Burst Switching   >10 μsec ~ 1 msec

Optical Packet Switching   > μsec

All-Optical Switching   > nsec

Feeder & Distribution Network Architecture

To backbone network

Ring 3

Ring 1

Ring 2

Ring 2
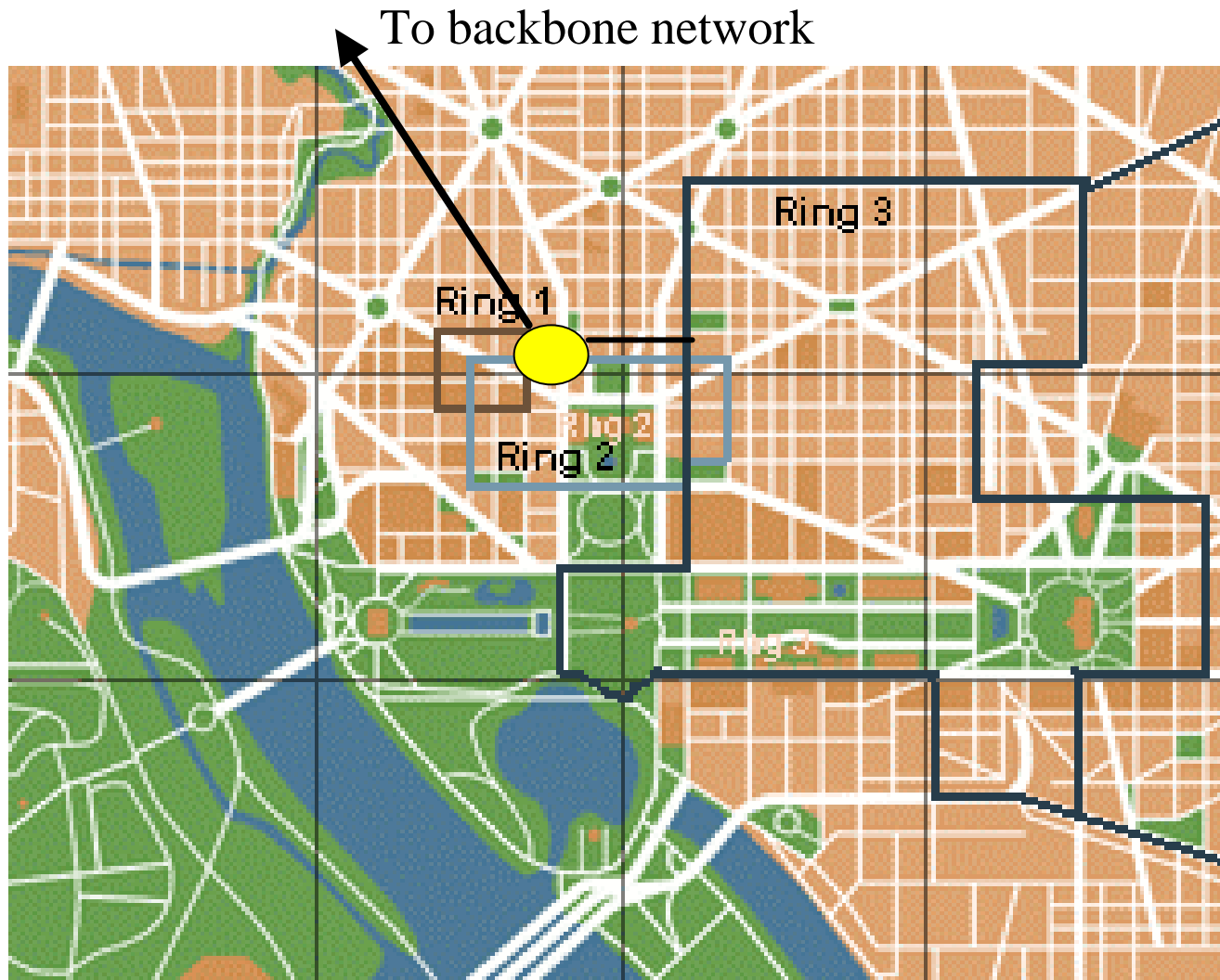
- service flexibility
- cost-effective architecture

ONRAMP
MIT, LL,AT&T,
JDS,Bay,Cabletron

ONRAMP Testbed

Backbone Network

POP

POP

AN = Access Node
EN = Egress Node

EN

AN/EN

Configurable Feeder Ring Network

AN

AN

AN

Band, Power, or Cable Splitter

Wavelength Add/Drop or Tap

Broadcasting Star or WGR

"Customers"
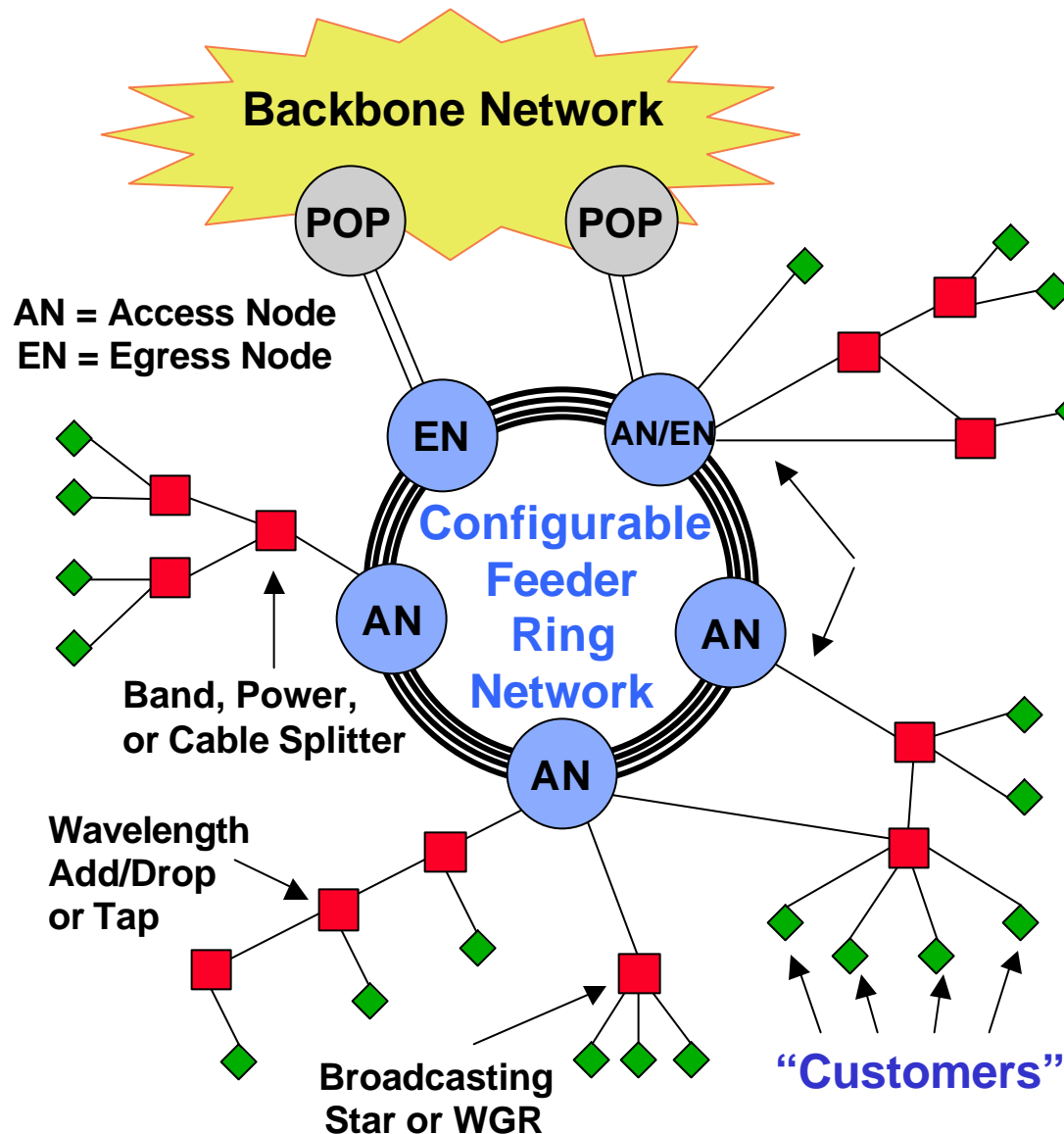
•Regional Access Network Architecture (10-1000 sq miles)

•Feeder Ring Network
   -multi-fiber WDM ring
   -reconfigurable Access Nodes
   -full optical restoration

•Distribution Network
   -cost sensitivity
   -passive, transparent WDM
   -tree/bus/ring topology

•BW squandering to mitigate complexity?
•wavelength density in feeder vs distribution network?
•shared or routed wavelengths?
•optical bypass, MAC protocol
•push end-node performance

Today's High Speed Access Solution

CO
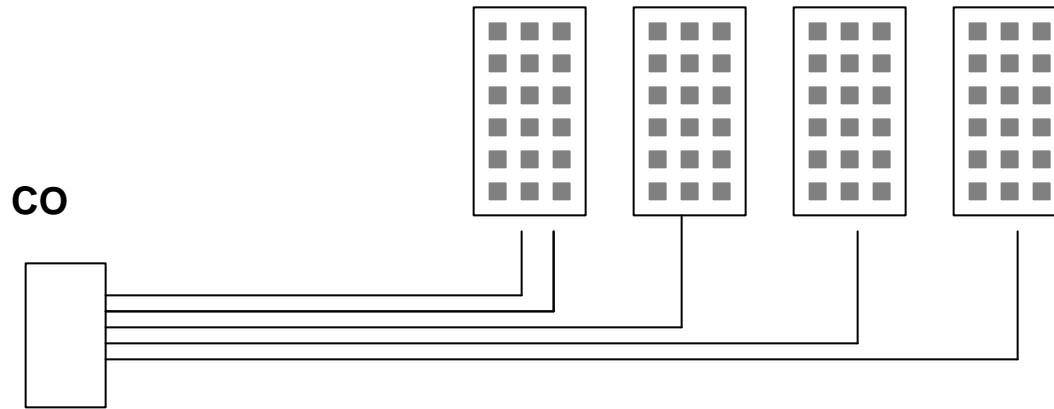
-individual fiber to end-customer sites requesting large bw
-multiplexing / switching in upstream office

A better solution needed that
-doesn't waste so much fibers
-enable very fast provisioning or very large bw on demand

Optical Access

CO
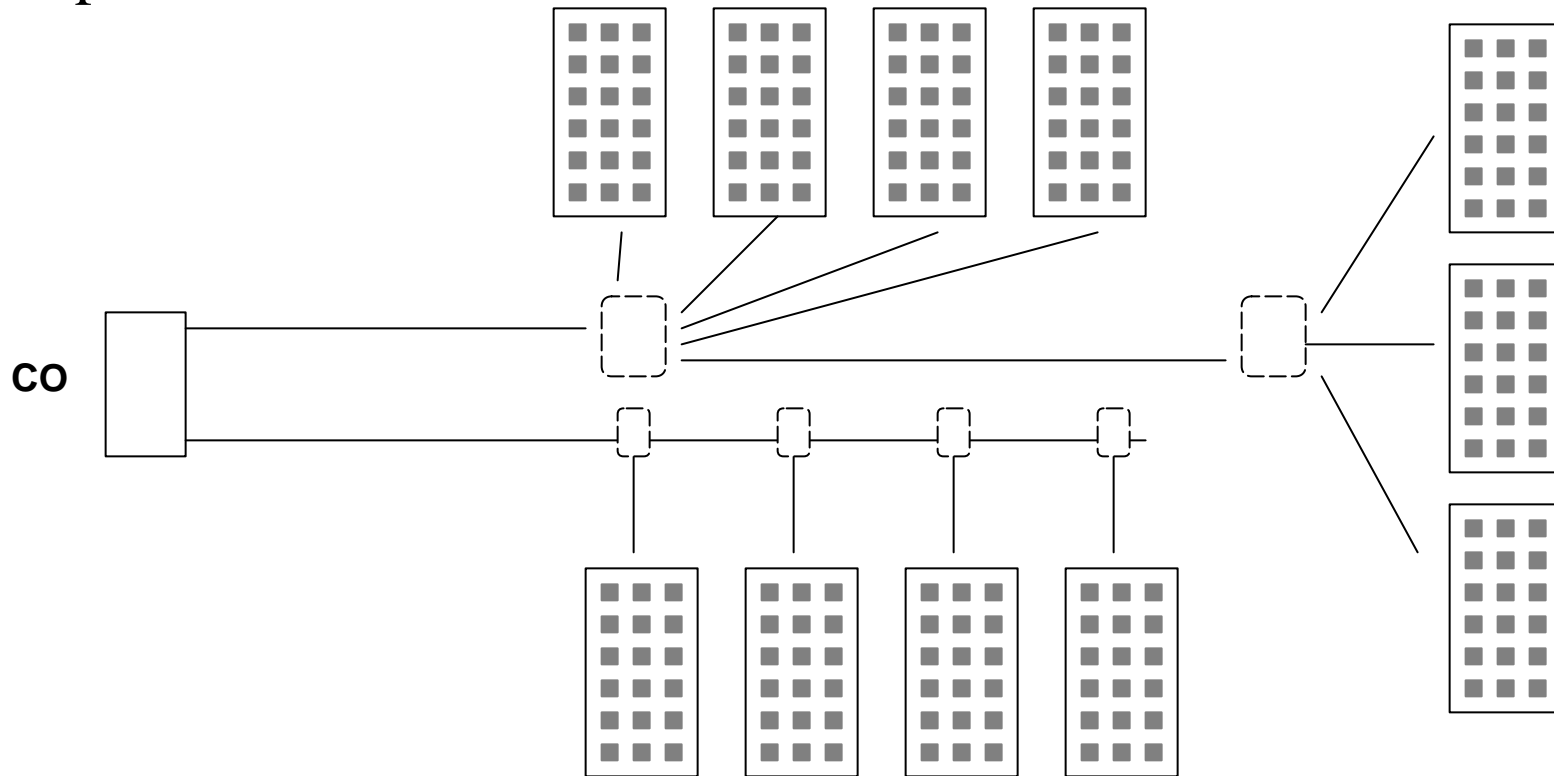
power
temperature control
environmental control

ECV
mini CO
manhole
bldg tel. room
pedestal

Need fiber infrastructure to be put in place - but use passive and active optical nodes in the distribution network

star, tree, bus, ring

Distribution of Routed Wavelengths

Passive Components in Distribution Network

Gigabit/second Host Platform

- Enable gigabits (up to 10 Gbps) to the end user
- Cell/Packet switch replaces traditional bus with its bottleneck
- Two new adapter cards being designed to plug into host switching backplane
- Network Service Card
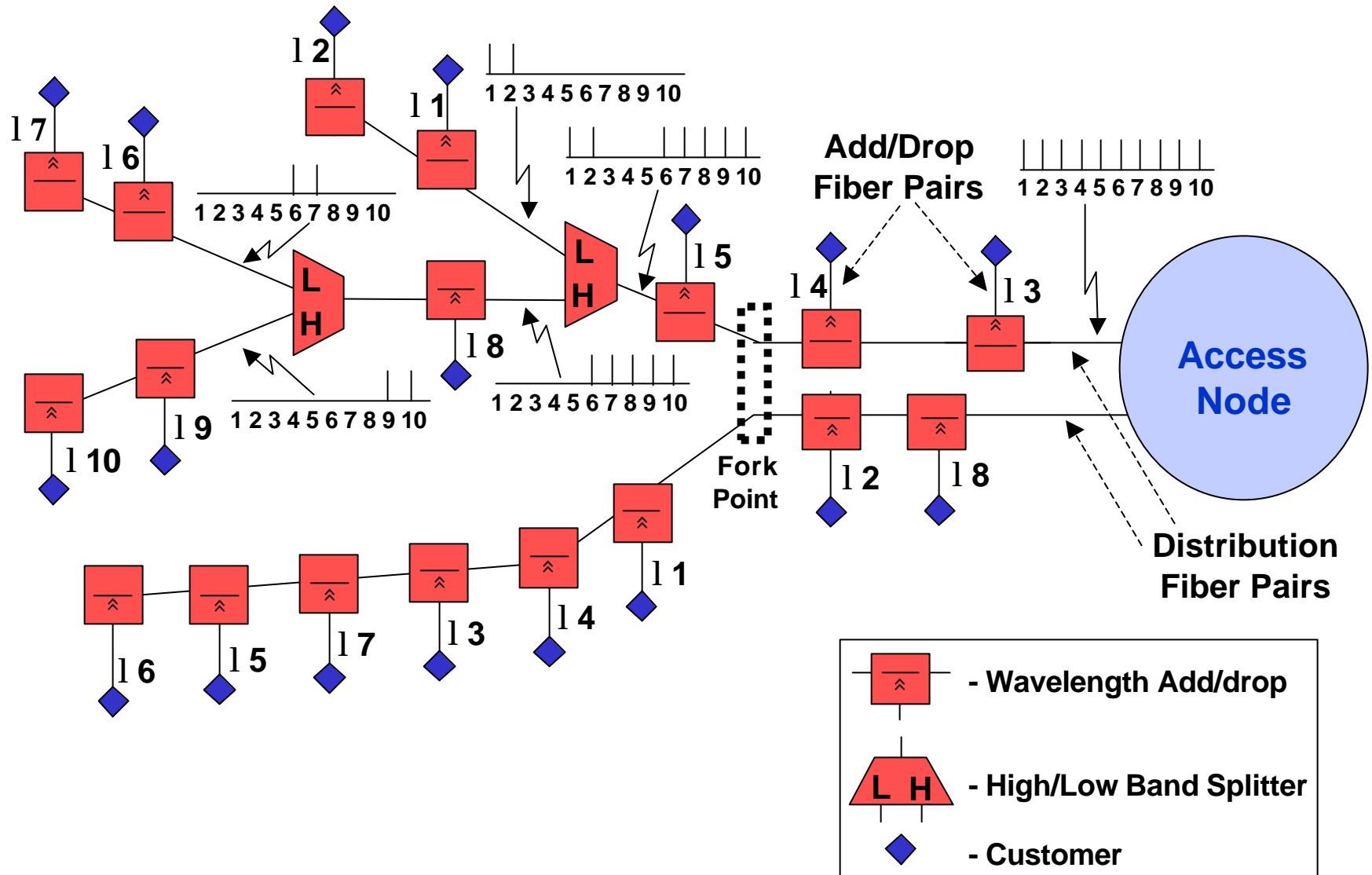   - offload many higher layer functions from host CPU, ATM & IP,
    - bursty & streams,
- Sensor Adaptor Card
    - multigigabit (bitrate agile)real-time stream from remote sensors to host
    for processing, storage, display

Traditional Host Computer Architecture

Cache | RAM | GPU | Display

CPU — Memory Ctrl — Bridge — PCI Bus

I/O Ctrl | I/O Ctrl

Ethernet | SCSI

Next Generation Architecture

GPU | Display

Cache | RAM

CPU — Memory Ctrl — Packet Switch — I/O Ctrl — e.g., IEEE 1394

NSC | SAC

ATM, IP on SONET | e.g. SMPTE 292

LAN, MAN, WAN | Remote Sensors
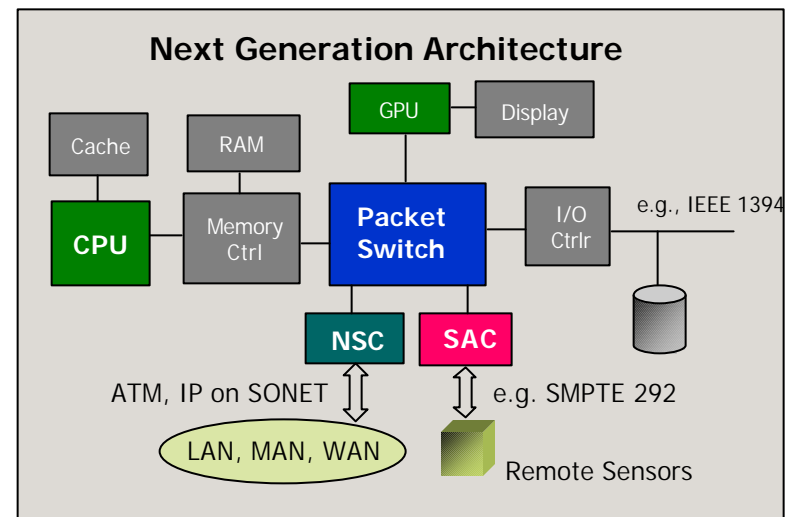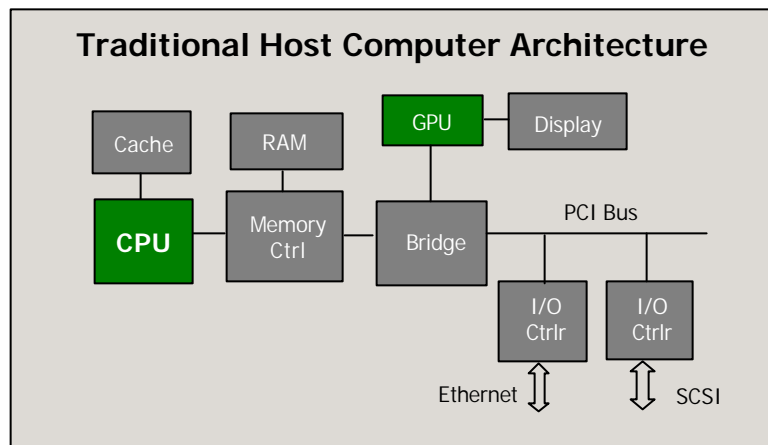
Network Elements Inc.

Multi-In Multi-Out (MIMO)

*random scattering decorrelates
multipath signals
multiple, independent parallel
channels*
capacity increases linearly with N!

**primitive
bit stream**

**1: N
demux**

S_1

S_2

S_N

**RANDOM HEAVY
SCATTERING**

1

2

M

**estimate/
decode**

Tx's

Rvr's

tradeoff: acquisition sequence overhead
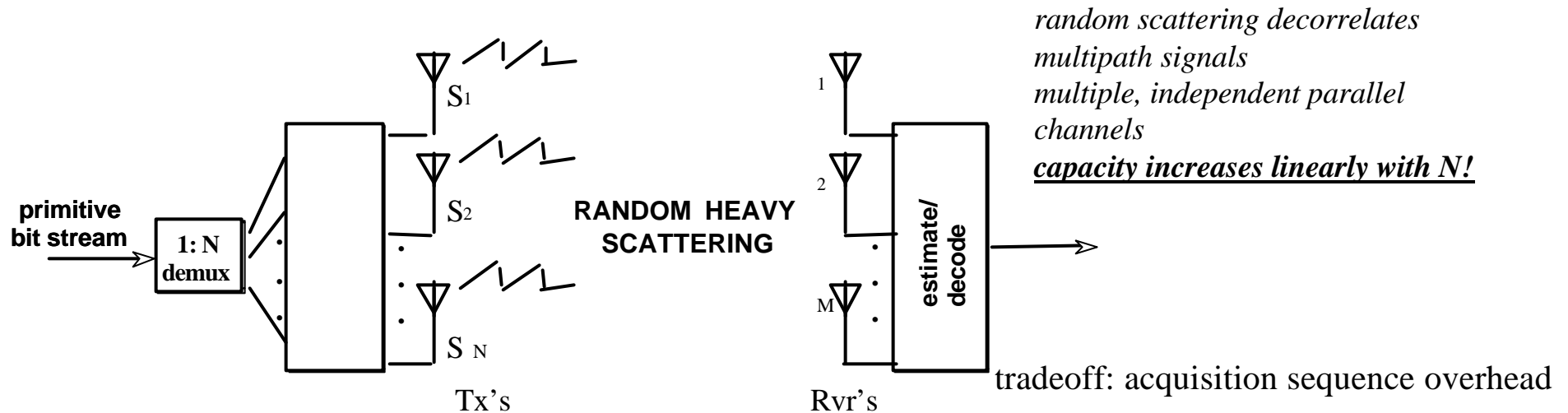
- Multiple antennas all transmitting in the same band
- Leverage heavy multipath environment
- Receiver signal processing:
  - Treat each sub-channel as "desired" signal, rest as "interferers" and use Adaptive Antenna Array-like
  technique to detect each (i.e. linear combinatorial nulling)
- Theoretically N-fold enhancement in spectral efficiency
- Acquisition time / training sequence overhead tradeoff

- *indoor conditions demonstrated: 20-40 bps/Hz over 30 kHz*
- *investigate higher bit-rates and outdoor (urban, suburbs, open terrain) system*

Networked Applications Performance Analysis: NetLogger Toolkit

application, host, network

•Application to application performance analysis tool
•Identifies bottlenecks in path of data flow: application ,
operating system, network level (e.g. CPU load, interrupt rate,
TCP retransmission, window size...)
•Post-hoc and real-time analysis
•Eveng Log Generation , Analysis and Visualization Tools
(depict event points, load-line, lifeline)

NetLogger/NLV analysis of a TerraVision with DPSS
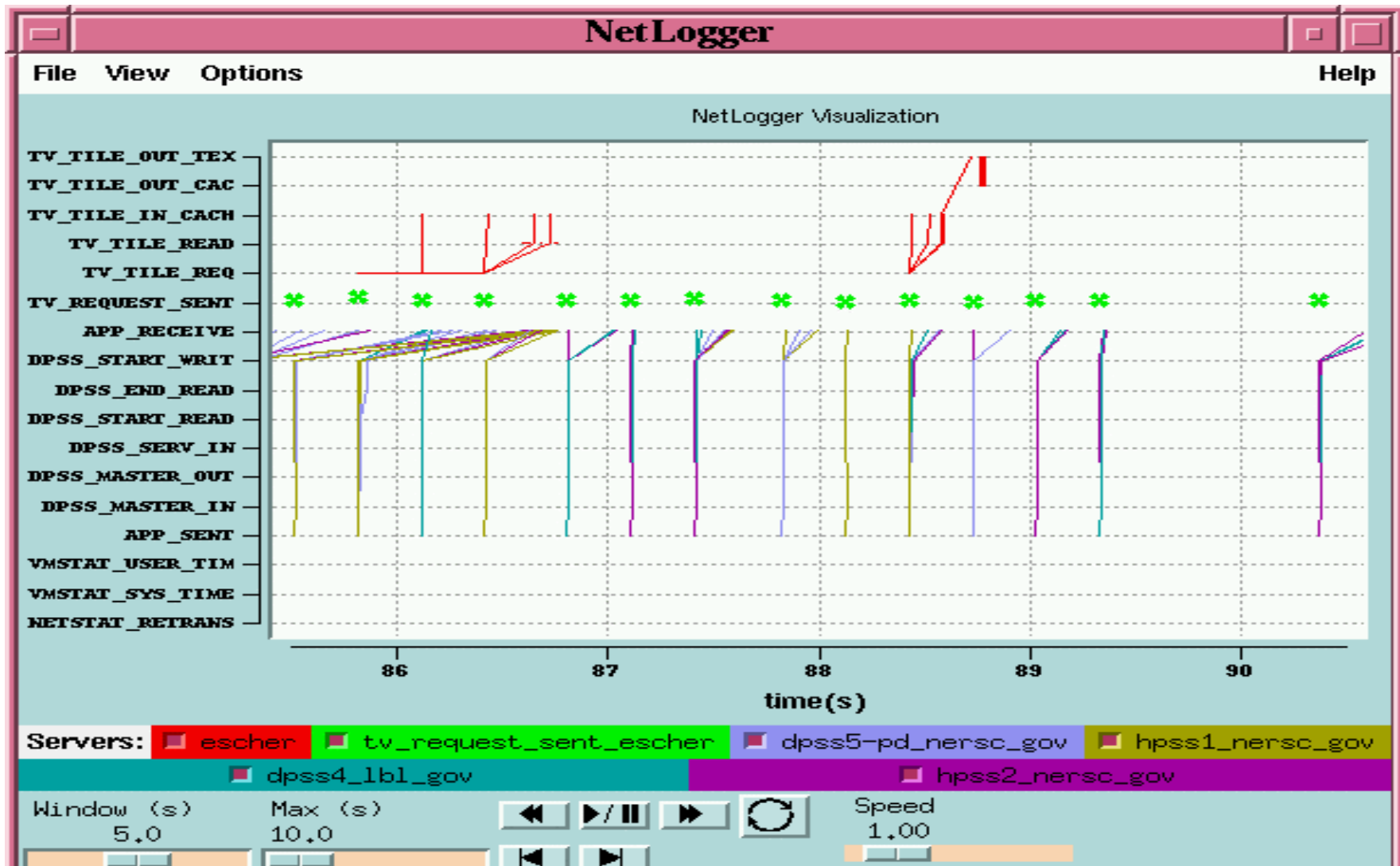
Network Engineering:
Network Monitoring, Analysis and Visualization

- Monitor and automate the discovery of the topology and traffic behavior of the Internet and future networks on a global scale.
- What makes this hard:

  –no central authority
  –scale (span and speed)
  –capturing dynamic behavior
  –visualization

Tools :
  "skitter" (active measurements: performance, topology)

  "coral" monitors (passive measurements over high speed links)

UCSD/CAIDA
(Cooperative Association for Internet Data Analysis)

Coral Tool: OC3mon

transit traffic
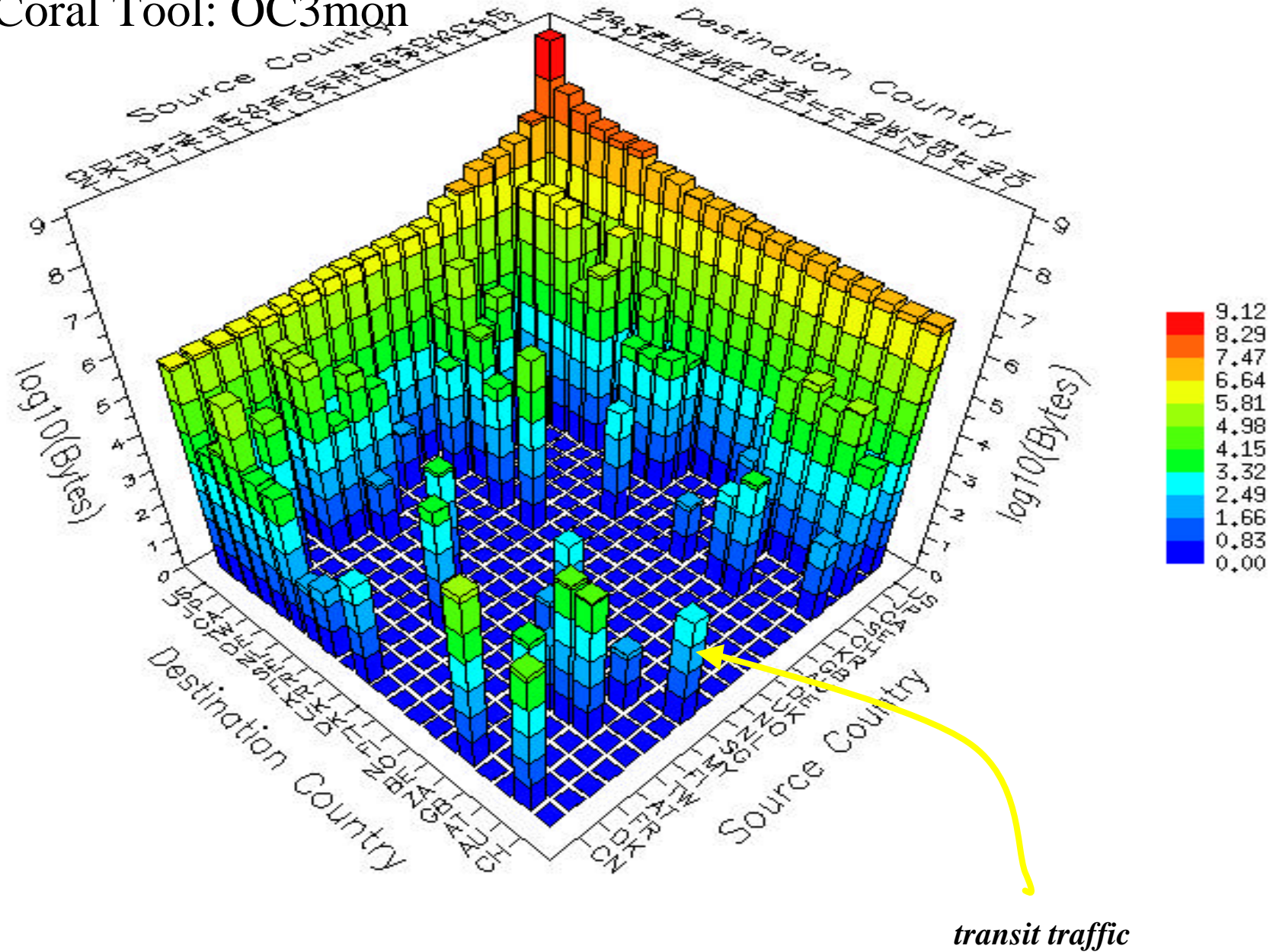
980312.1532UT.bytes_bycountry
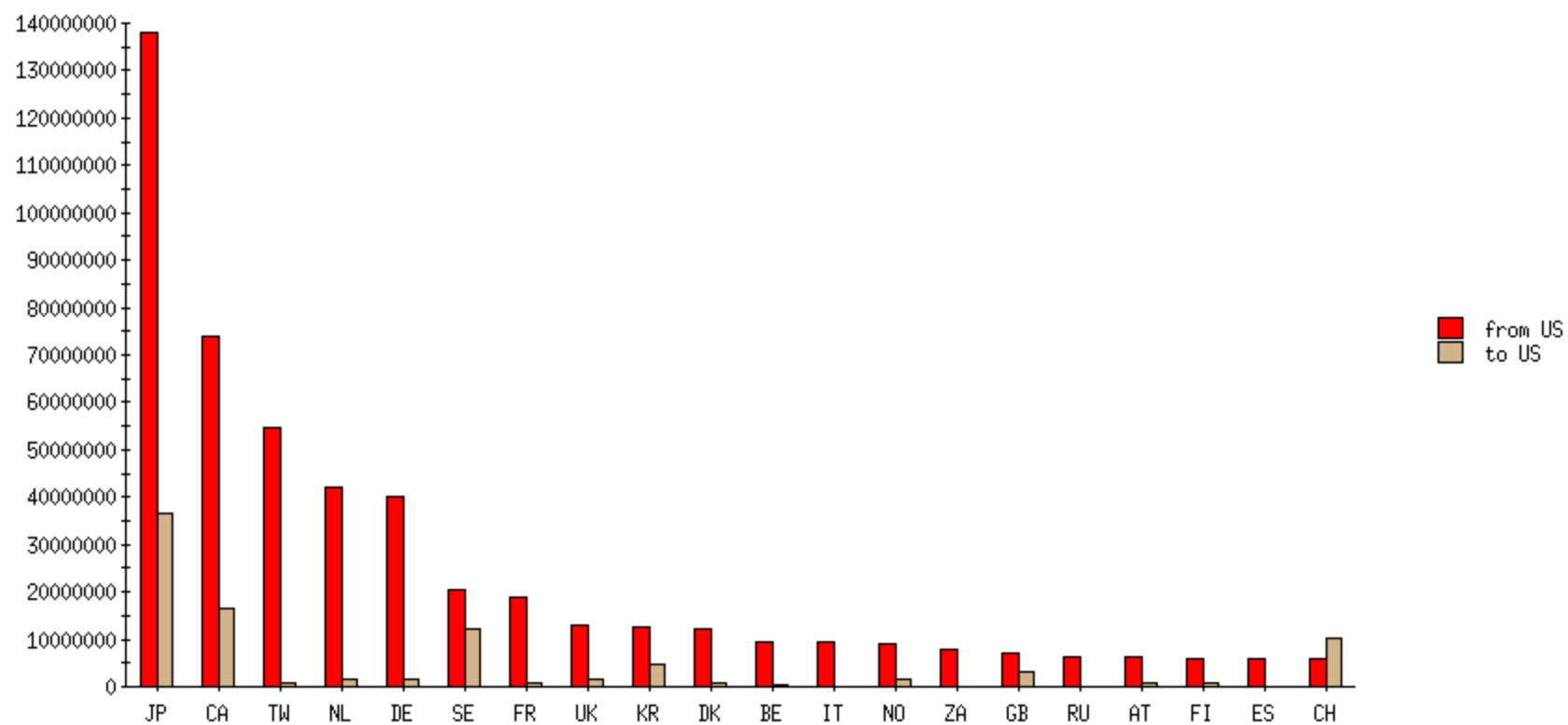
Network Tomography

- Network "Radar" :Global connectivity information
- Measure IP paths ("hops") from source to MANY (~10^4) destinations
- Use 52 byte ICMP echo requests (every 30 min.) as probes
- Challenges:
  - pervasive measurement with minimal load on infrastructure
  - visualization

Internet Tomography

hop count histogram

temporal behavior

GTrace v1.0.0

Trace Maps Tools Options Help

Host/URL to Trace www.utah.edu ▼ Lat/Lon :

Location Legend

- 🟢 LOC Record
- 🟡 Complete DB/TextFile Match
- 🔵 Domain Parsing File
- 🔵 Hostname Reduction Match
- 🔴 whois Record
- ⚪ Country Code

Ok

Location Details

Location obtained from rules in domain parsing file

Latitude   : 34.05
Longitude   : −118.24
City   : Los Angeles
State   : CA
Country   : US

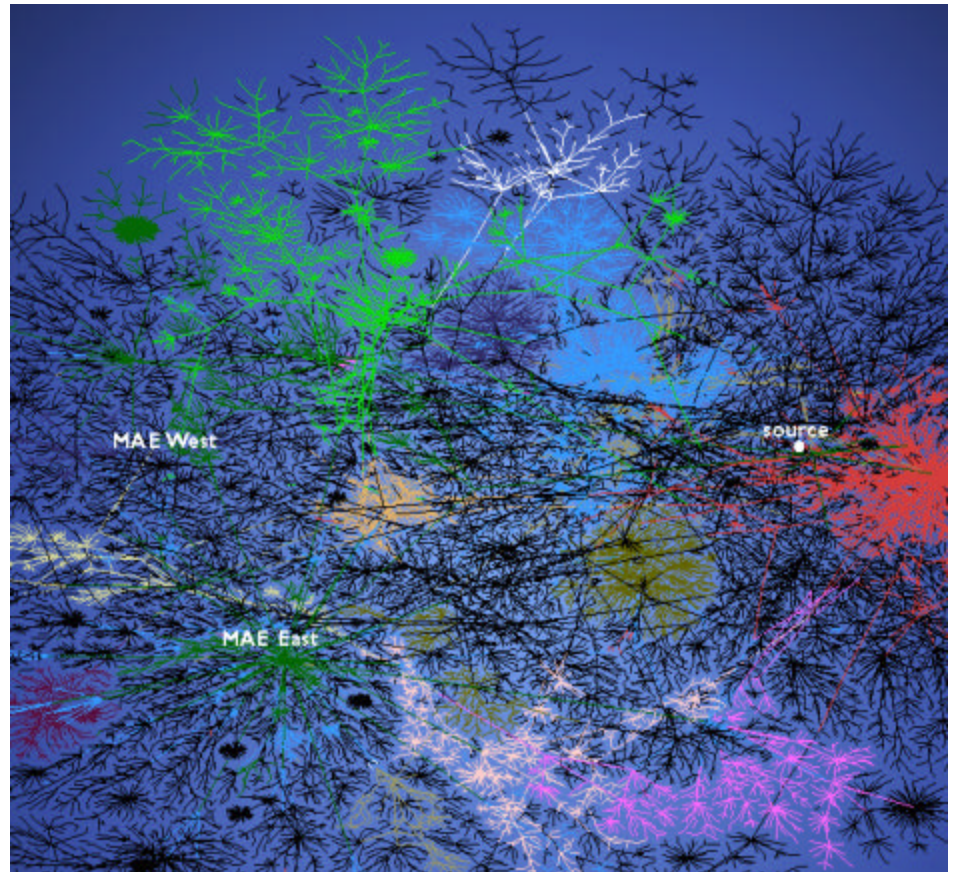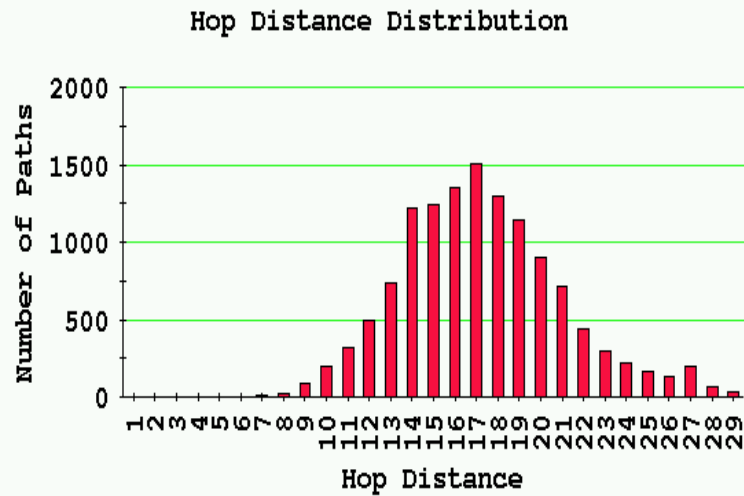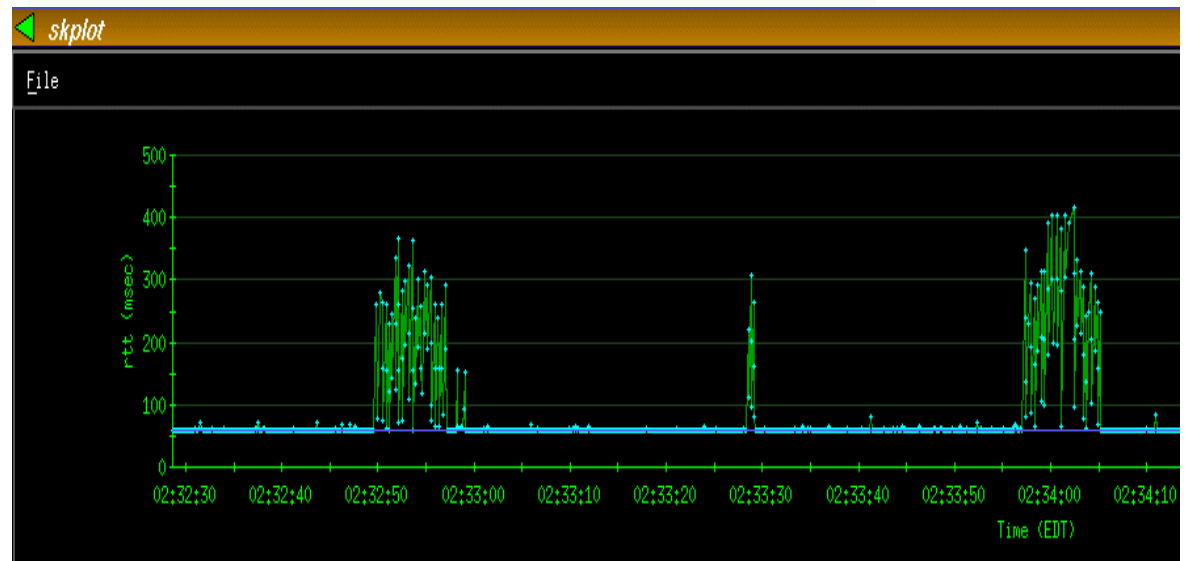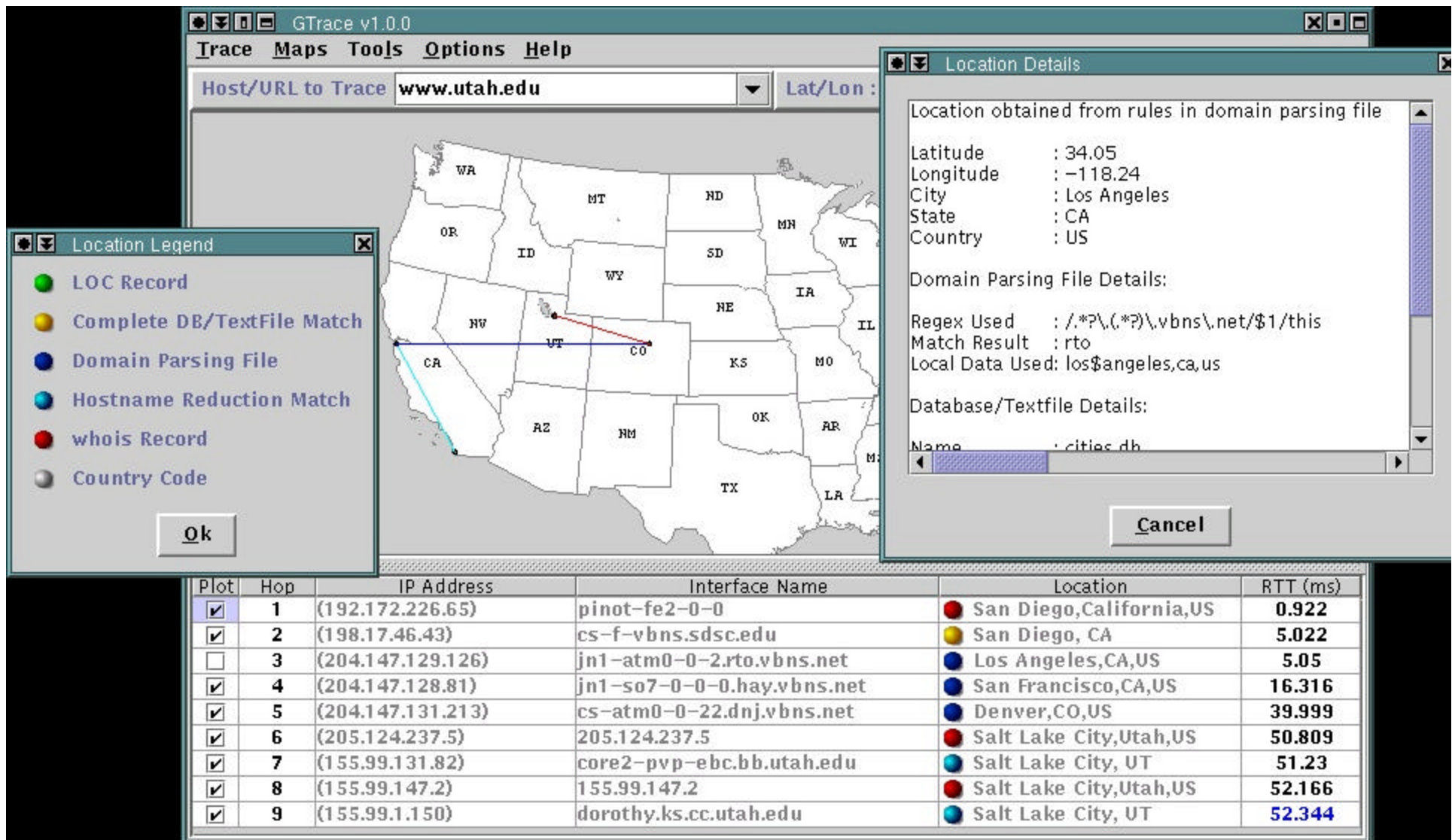Domain Parsing File Details:

Regex Used   : /.*?\.(.*?)\.vbns\.net/$1/this
Match Result : rto
Local Data Used: los$angeles,ca,us

Database/Textfile Details:

Name   : cities.db

Cancel

Plot	Hop	IP Address	Interface Name	Location	RTT (ms)
☑	1	(192.172.226.65)	pinot-fe2-0-0	🔴 San Diego,California,US	0.922
☑	2	(198.17.46.43)	cs-f-vbns.sdsc.edu	🟡 San Diego, CA	5.022
☐	3	(204.147.129.126)	jn1-atm0-0-2.rto.vbns.net	🔵 Los Angeles,CA,US	5.05
☑	4	(204.147.128.81)	jn1-so7-0-0-0.hay.vbns.net	🔵 San Francisco,CA,US	16.316
☑	5	(204.147.131.213)	cs-atm0-0-22.dnj.vbns.net	🔵 Denver,CO,US	39.999
☑	6	(205.124.237.5)	205.124.237.5	🔴 Salt Lake City,Utah,US	50.809
☑	7	(155.99.131.82)	core2-pvp-ebc.bb.utah.edu	🔵 Salt Lake City, UT	51.23
☑	8	(155.99.147.2)	155.99.147.2	🔴 Salt Lake City,Utah,US	52.166
☑	9	(155.99.1.150)	dorothy.ks.cc.utah.edu	🔵 Salt Lake City, UT	52.344

Scaling the number of Flows

OC3 Link
60,000~100,000 flows over
5 minute period
timeout after 1 minute

Flow-Size Histogram
(March 1999)

Fine-Grained
Networking

SuperNet
Technology

Number of flows

20K

15K

10K

5K

0

10b   100b   1Kb   10Kb   100Kb   1Mb   10Mb   100Mb

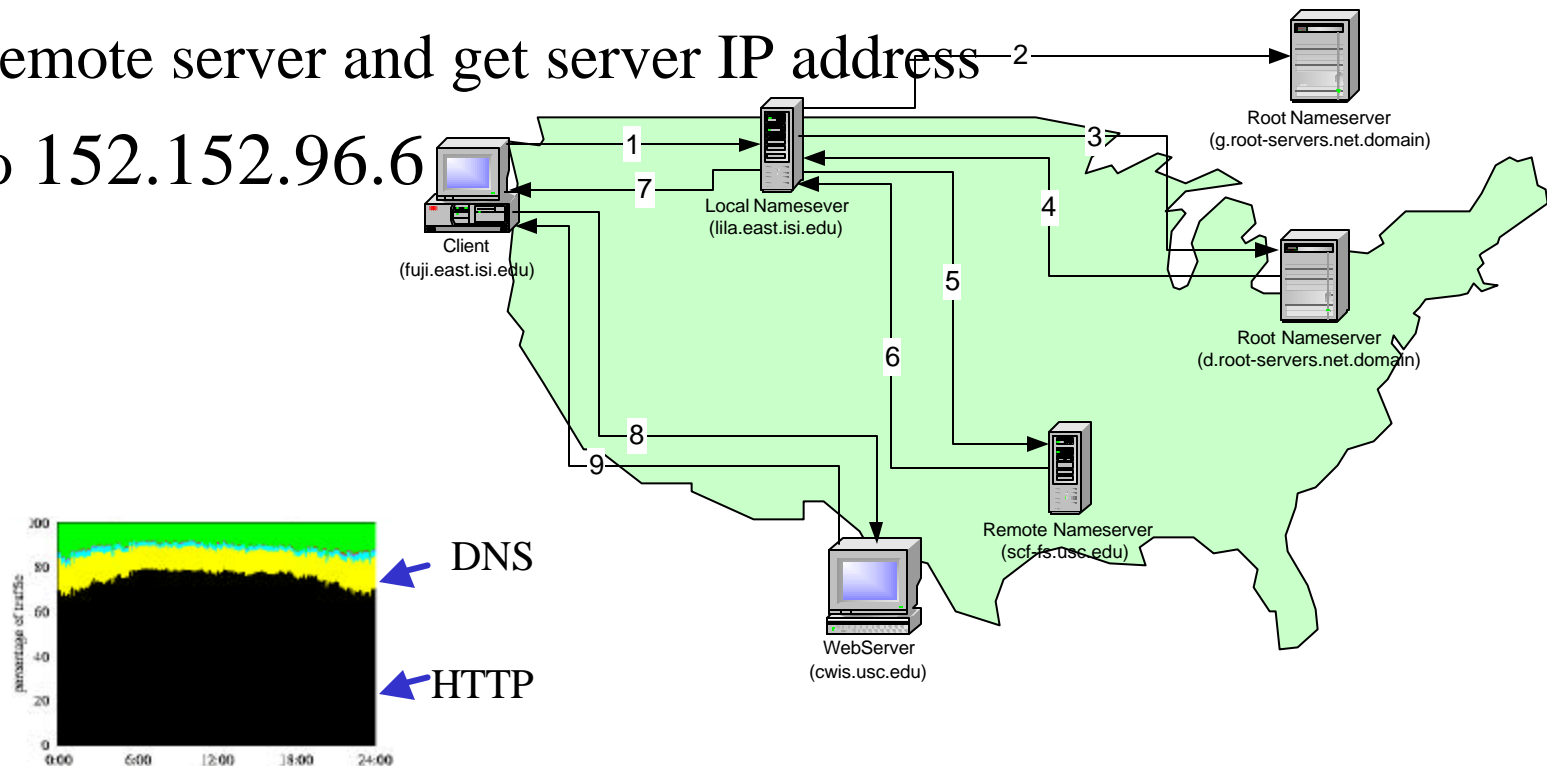Flow Size (Bytes)
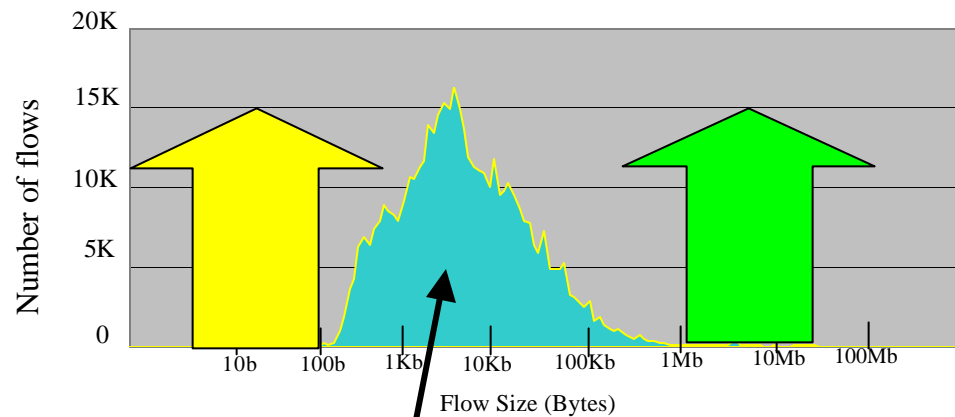
web-centric traffic

Name Lookup Today

- local lookup
  "http://www.nato.int/kosovo/video.htm"

- if not cached, go to root name servers & get remote
  DNS IP address

- go to remote server and get server IP address

- send to 152.152.96.6

2

Root Nameserver
(g.root-servers.net.domain)

1

3

7

Local Nameserver
(lila.east.isi.edu)

4

Client
(fuji.east.isi.edu)

5

Root Nameserver
(d.root-servers.net.domain)

6

8

9

Remote Nameserver
(scf-fs.usc.edu)

DNS

HTTP

WebServer
(cwis.usc.edu)

percentage of traffic

0:00   6:00   12:00   18:00   24:00

# Next-generation networking and service environment

- Network with devices & sensors, plus computers
- Devices, users (computers) and services may be mobile
- Services may be composed of groups of nodes
- Problems: configuration, routing, discovery, adaptation, security

App should be able to conveniently (i) specify a resource and (ii) send messages to it

iNAT Project (MIT)

- Intelligent naming
  - Intentional Naming System (INS)
  - Resource discovery in future networks
- Adaptive transmission
  - Congestion Manager (CM)
  - End-system congestion management and adaptation in the future Internet

InterPlanetary Networking

venus.ipn

moon.ipn

earth

mars.ipn

jupiter.ipn
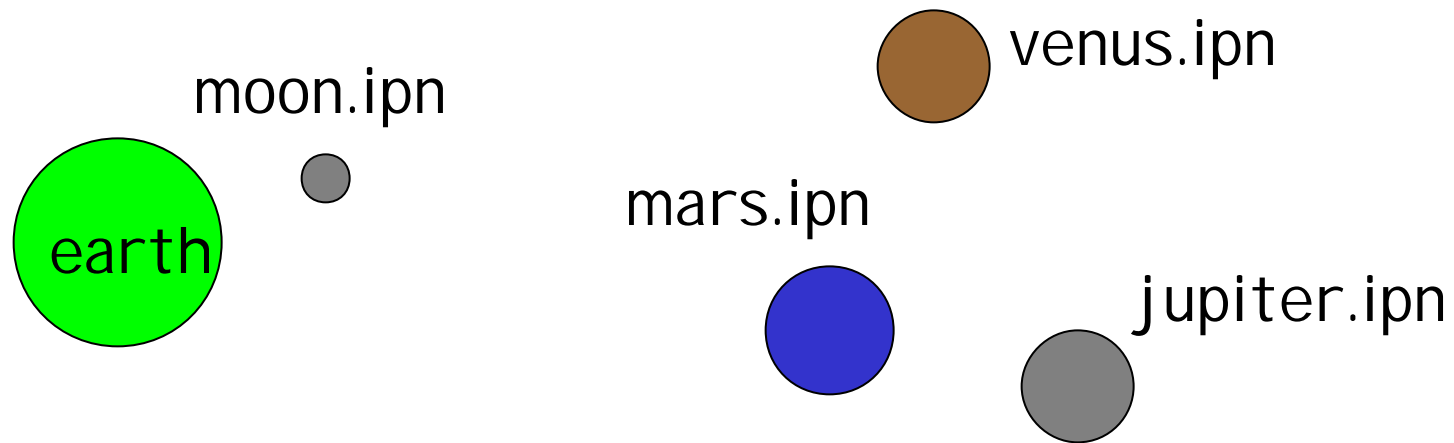
- time-dependent, high latency, lossy paths in deep space
- between planetary gateways, internets, platforms
- layer 2, 3, 4 protocols
- IP address space and naming in space (domain name server)

> 100 Gb/s All-Optical Logic and Time-Division Multi-Access Network
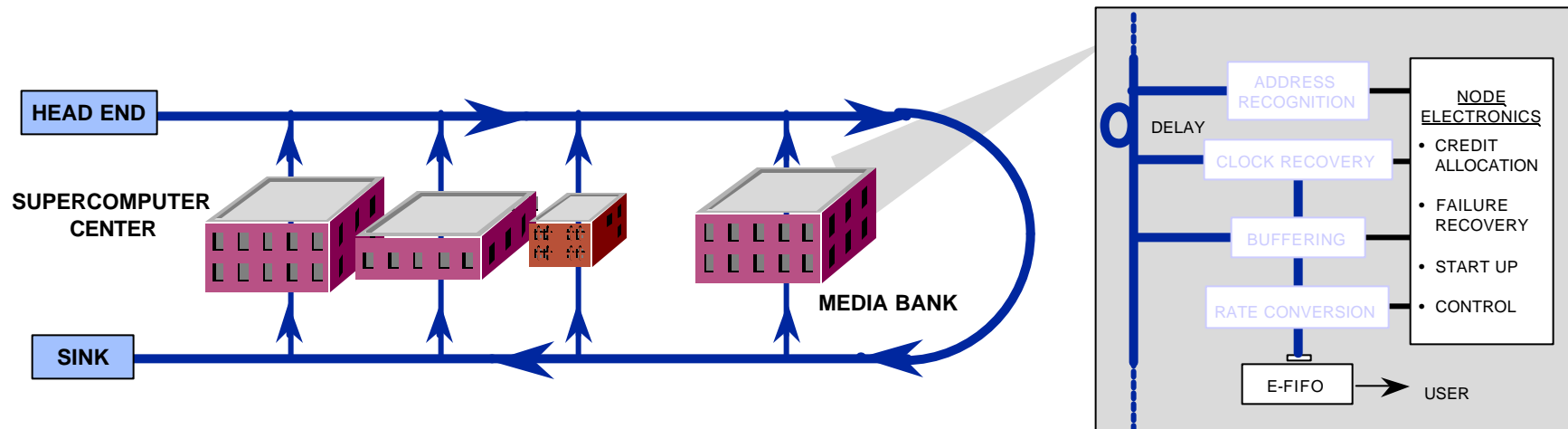
- ultra-fast all-optical logic
- instead of demultiplexing hierarchically into lower rates, enable users to seize 100Gbps+ stream
- implementation of network nodes and TDM-based LAN
- performer: MIT Lincoln Laboratory

All-Optical Logic Gates
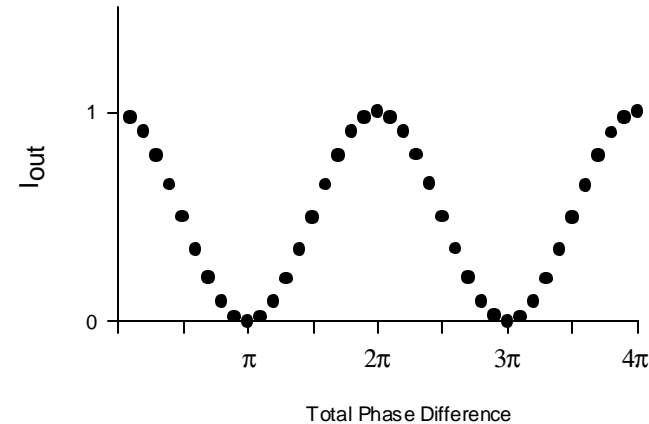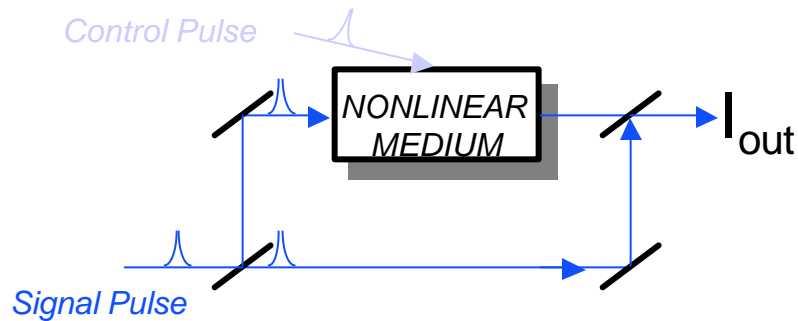
- nonlinear Kerr effect induces a phase shift in presence of control pulse
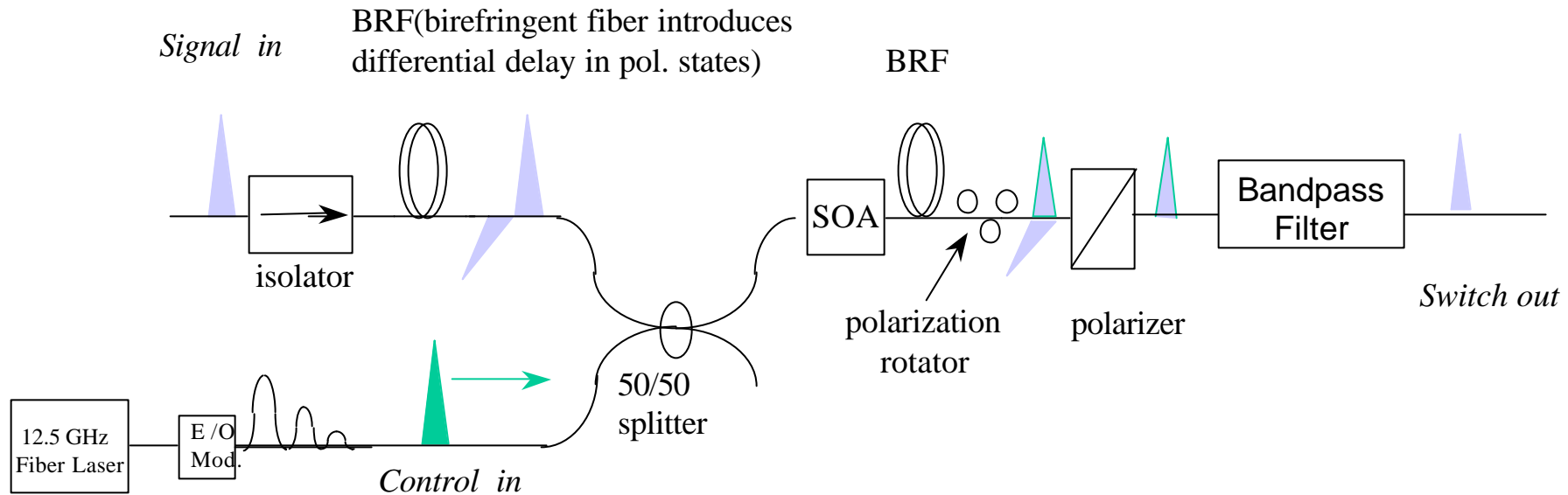
$$I_{out} \sim \cos^2((\phi_b + \Delta\phi_{nl})/2)$$

$$n = n_0 + n_2 I_c$$

$$\Delta\phi_{nl} = (2\pi/\lambda)\, n_2\, L\, I_c$$

- nonlinear element : fiber or SOA (semiconductor optical amplifier)
- different configurations - TOAD, NOLM, UNI

All-Optical Switch Implementation

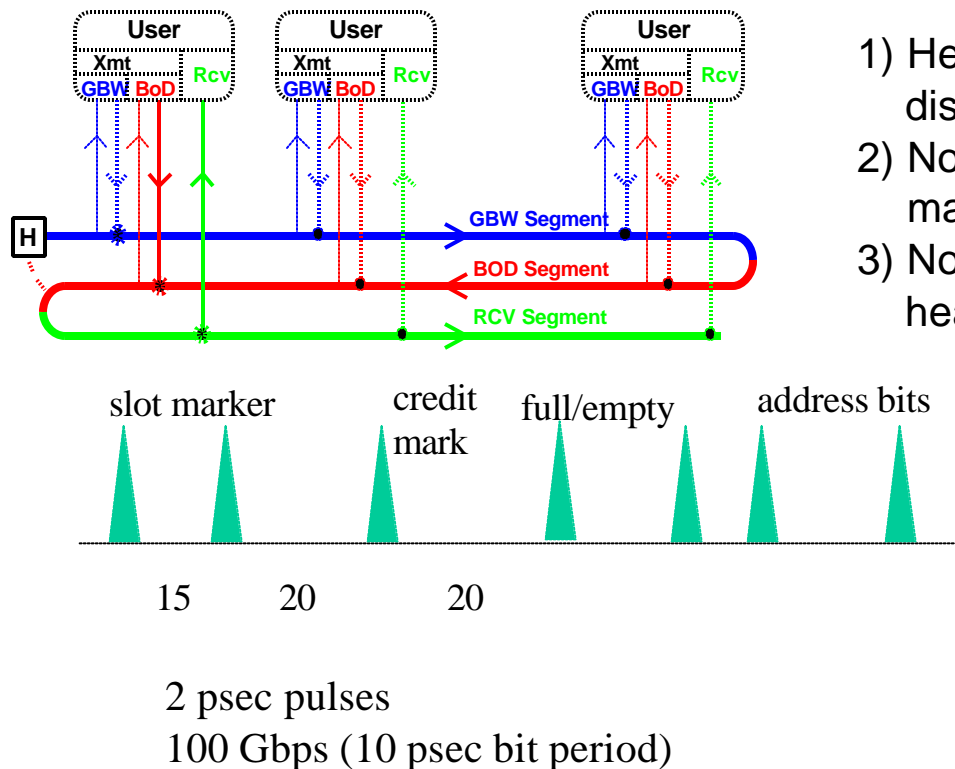## *Ultrafast Nonlinear Interferometer (UNI)*

- Ultra short pulse train sequence generated via split/delay/combine technique
- Ultrafast refractive index effects yield differential phase shift between two orthogonal signal
- components.
- Differential phase shifts translate into polarization changes of the temporally realigned signal
- components (pi shift translates to polarization rotation)

AND Operation: UNI biased OFF, control pulses turn signal ON

INVERT Operation: UNI biased ON, control pulses turn signal OFF
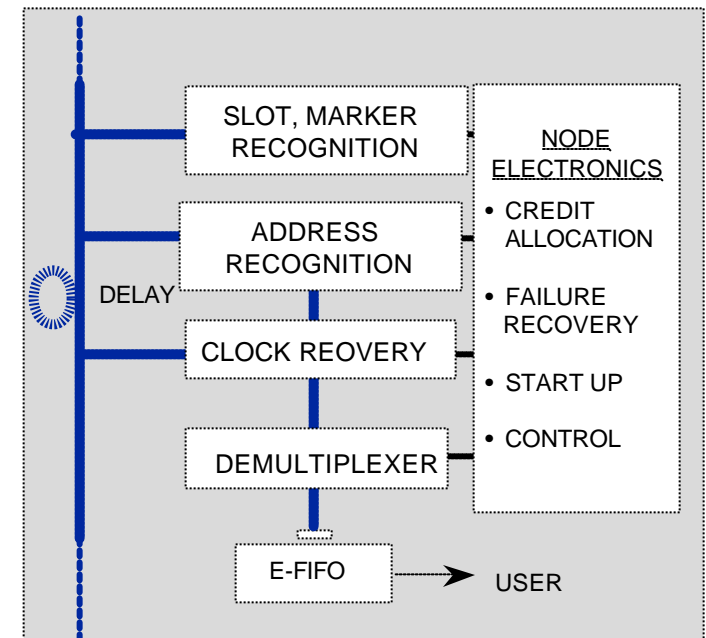
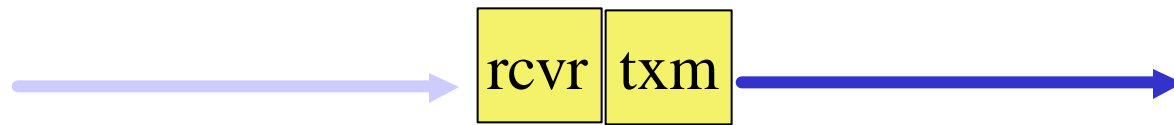# Networking Architecture and Protocol

Folded Unidirectional Bus

User
Xmt   Rcv
GBW BoD

User
Xmt   Rcv
GBW BoD

User
Xmt   Rcv
GBW BoD

GBW Segment

BOD Segment

RCV Segment

H

slot marker   credit   full/empty   address bits
                     mark

15   20   20

2 psec pulses
100 Gbps (10 psec bit period)

PROTOCOL

1) Head end generates empty slots and
   distributes credits
2) Nodes with data to send and a credit
   may access any empty slot on the network
3) Nodes receive packets from the bus, read the
   headers and process data intended for them

SLOT, MARKER
RECOGNITION

NODE
ELECTRONICS

• CREDIT
  ALLOCATION

ADDRESS
RECOGNITION

• FAILURE
  RECOVERY

DELAY

CLOCK REOVERY

• START UP

• CONTROL

DEMULTIPLEXER
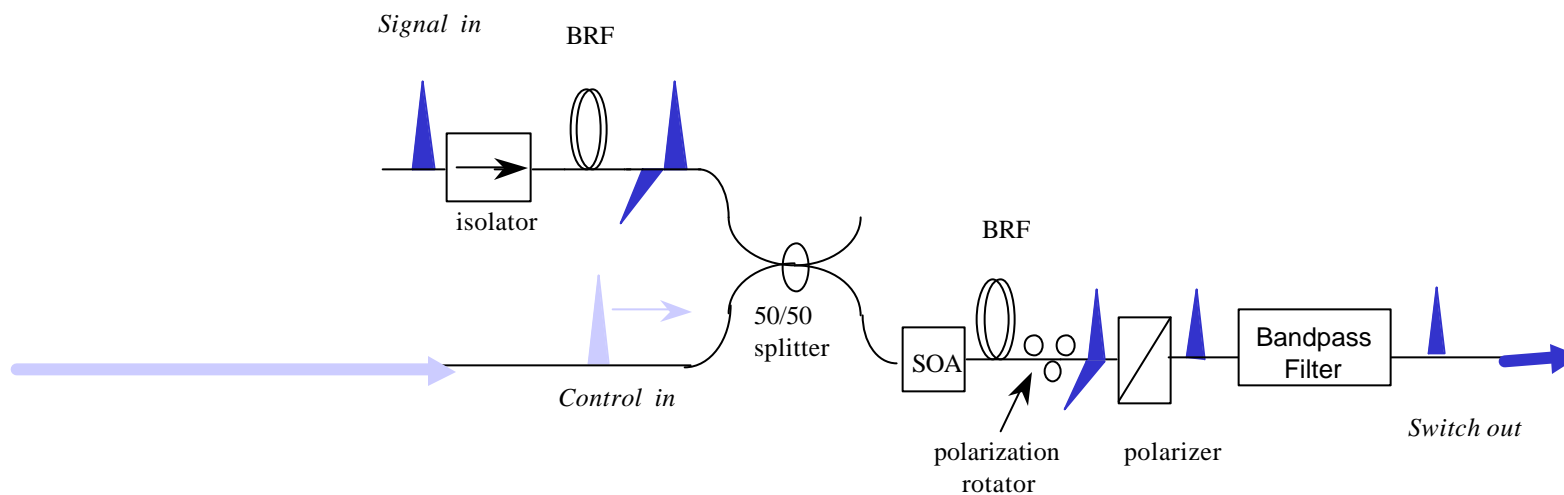
E-FIFO   USER

User Interface Nodes

RECEIVER

Today's Electro-Optic Regeneration

rcvr txm

electro-optic conversion
re-timing, re-shaping,
amplification

All Optical Regeneration

Signal in

BRF

isolator

50/50
splitter

Control in

BRF

SOA

polarization
rotator

polarizer

Bandpass
Filter

Switch out

SUPERNET TESTBED (www.ngi-supernet.org)

NTON II
4 λ's
@ 10 Gb/s per λ

Nortel

ONRAMP
Testbed

MIT

TCG

Cabletron

MIT Lincoln

AT&T

Seattle

Portland

Boston

BossNet
dark fibers

HSCC
2.5 Gbps

D.C.

dark fibers
in Bay Area

Bay Area

Los Angeles

San Diego

ISI-East

NSA

NASA

NRL

DISA

DIA

UC
Berkeley

GST

LLNL

LBNL

BART

Sprint

vBNS

SNL

SRI

NASA/
Ames

ATDNet / MONET

multi-vendor (opt,gigE, layer2,layer3)
multi-service provider
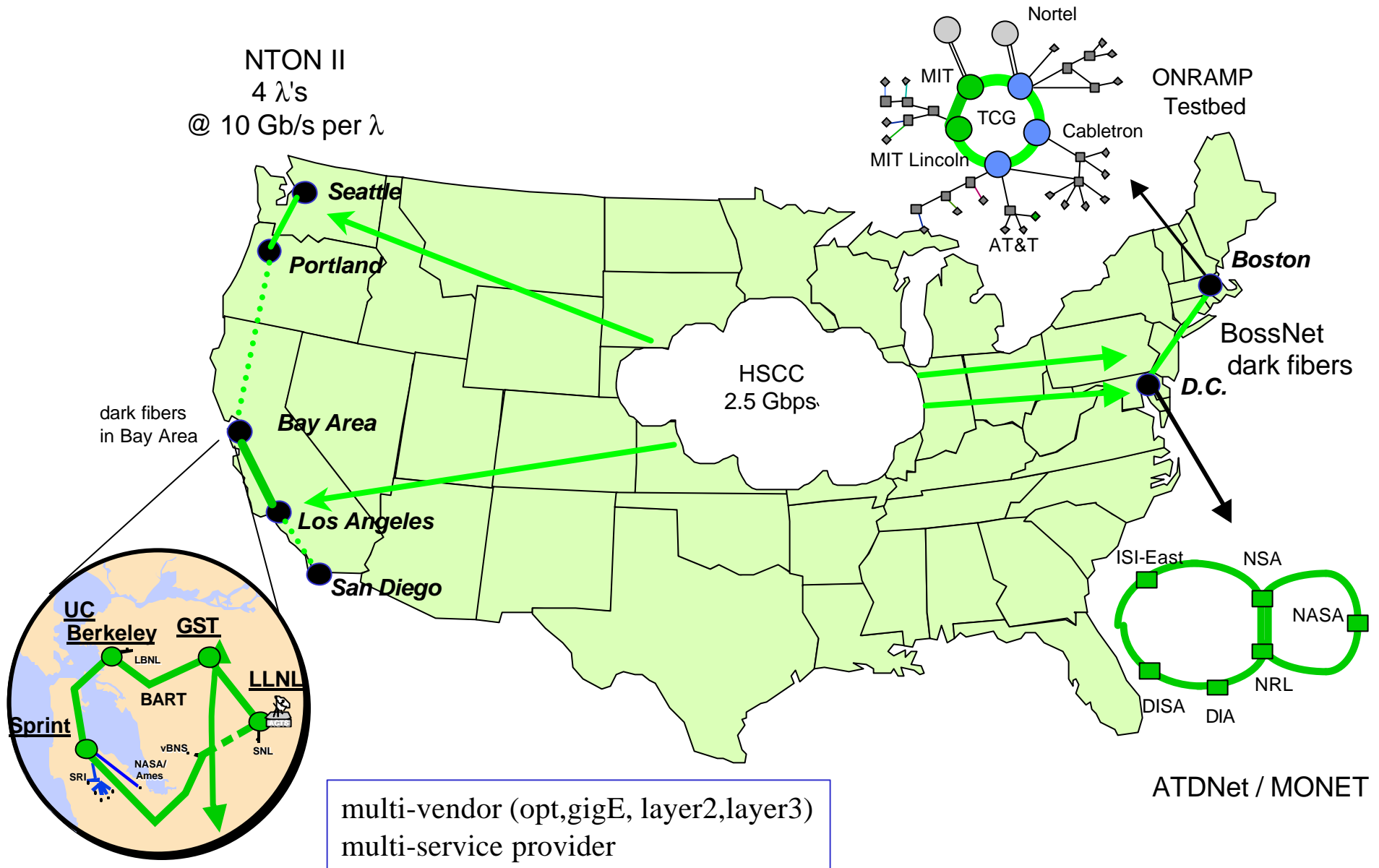
Sites Supporting 1 Gigabit+ Connectivity

- Connected Today
- Planned or under discussions for 2000
- {connected in at lower bw}

ONRAMP

- MIT Campus
- MIT Lincoln Laboratory
- Cabletron
- AT&T/TCG
- Harvard

NTON

- JPL
- CalTech
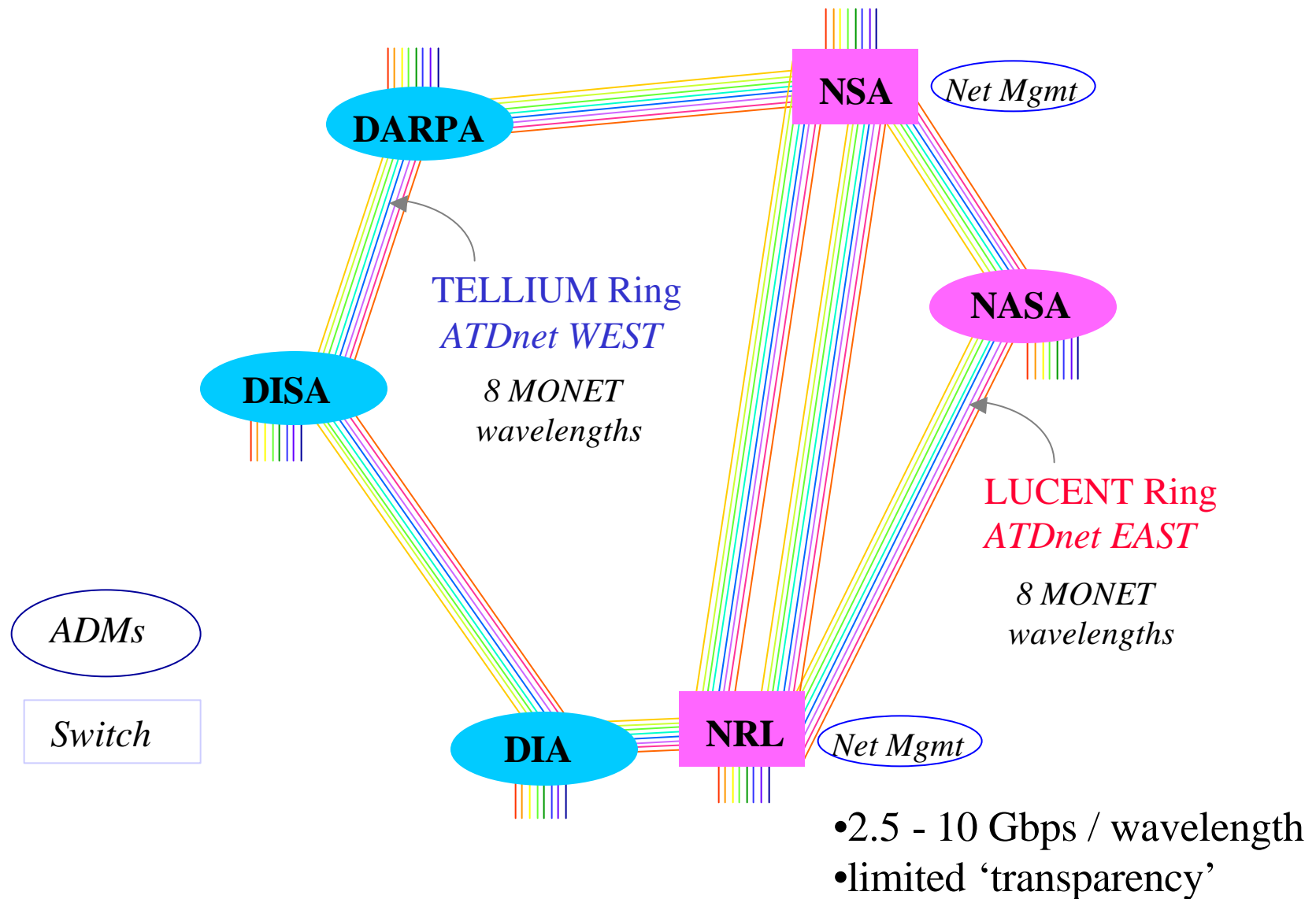-
- LLNL
- Sandia NL
- SLAC
- NASA Ames
- USC - almost
- Boeing
- Tektronix
- SDSC
- Spawar
- Network Elements Inc.
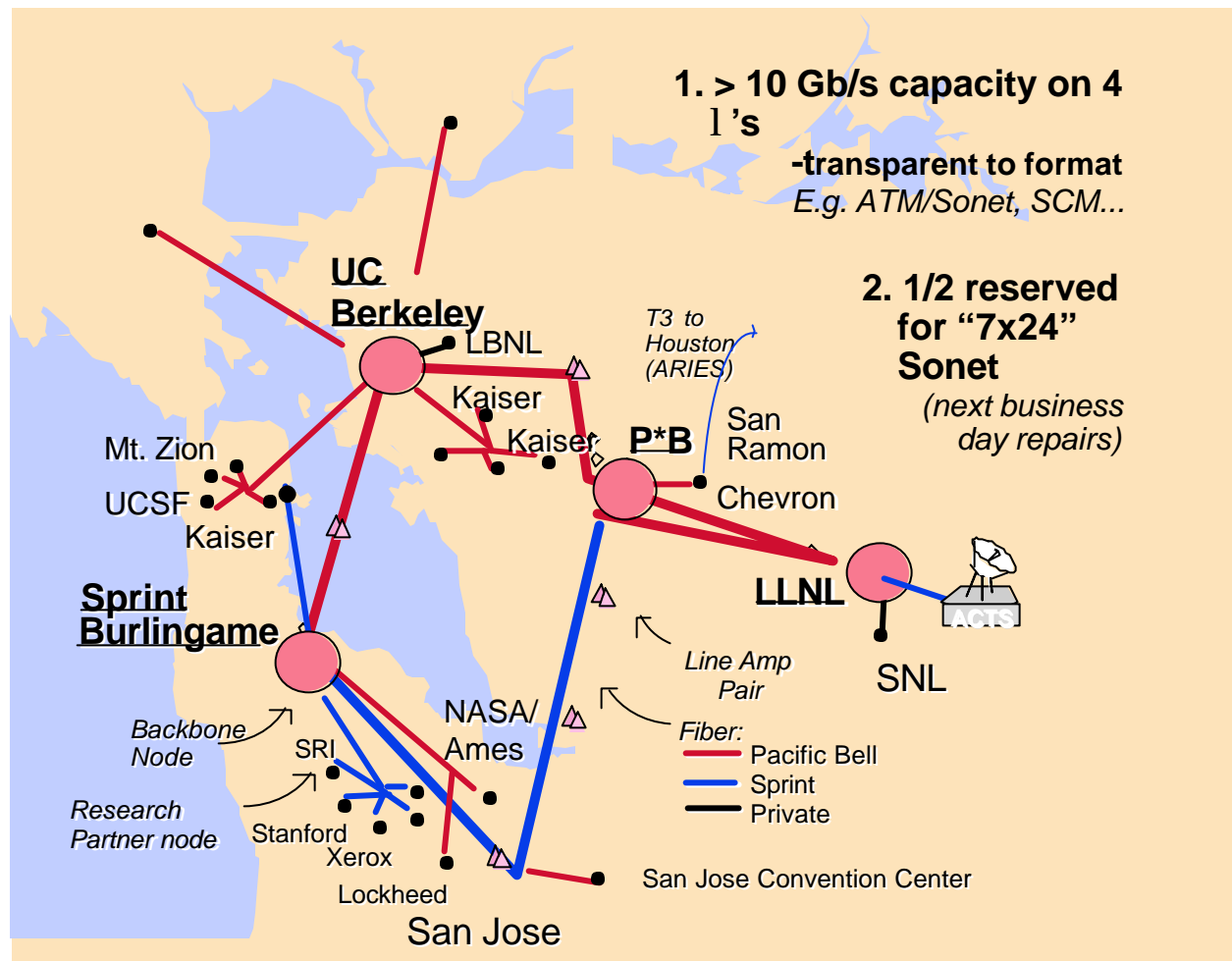- {Silicon Valley Test Track - Sprint, sun, sgi, xerox park ..}

HSCC

- U of Washington
- Microsoft
- CMU
- U Pitt Medical Center
- Pittsburgh SuperComputing
- Colorado State U
- NYC

BOSSNET

- Drexel
- U of Penn
- Sarnoff Laboratory
- {Johns Hopkins Medical School}
- {UMDNJ}

ATDNET/ MONET

- ISI -East/DARPA
- NRL
- NASA
- NSA-U of MD
- DISA
- DIA
- CNRI
- North Carolina -MCNC, UNC .
- {Walter Reed Army Hospital}
- {NIST}
- {NIH/NLM}
- {NIMA}
- {Naval Surface Warfare}
- {Holocaust Museum}
- {Office of Naval Intelligence}

ATDNET-MONET TESTBED

NSA

Net Mgmt

DARPA

TELLIUM Ring
ATDnet WEST

*8 MONET
wavelengths*

NASA

DISA

LUCENT Ring
ATDnet EAST

*8 MONET
wavelengths*

ADMs

Switch

DIA

NRL

Net Mgmt

- 2.5 - 10 Gbps / wavelength
- limited 'transparency'

Research Partner Application Sites

1. > 10 Gb/s capacity on 4 λ's
   -transparent to format
   E.g. ATM/Sonet, SCM...

2. 1/2 reserved for "7x24" Sonet
   (next business day repairs)

UC Berkeley
LBNL
Kaiser
Kaiser
P*B
San Ramon
Chevron
T3 to Houston (ARIES)
Mt. Zion
UCSF
Kaiser
Sprint Burlingame
Backbone Node
Research Partner node
SRI
NASA/ Ames
Stanford
Xerox
Lockheed
San Jose
LLNL
SNL
ACTS
Line Amp Pair

Fiber:
— Pacific Bell
— Sprint
— Private

San Jose Convention Center

BOSSNET Testbed

- Four fibers along inland/coast rail routes between Washington DC and Boston
- 29 huts being populated with custom equipment (span length 40-100km)
- Connection between HSCC, MONET/ATDNet, ONRAMP networks

Boston — ONRAMP

two fibers →

NYC

four fibers →

DC

HSCC   MONET/ATDNET
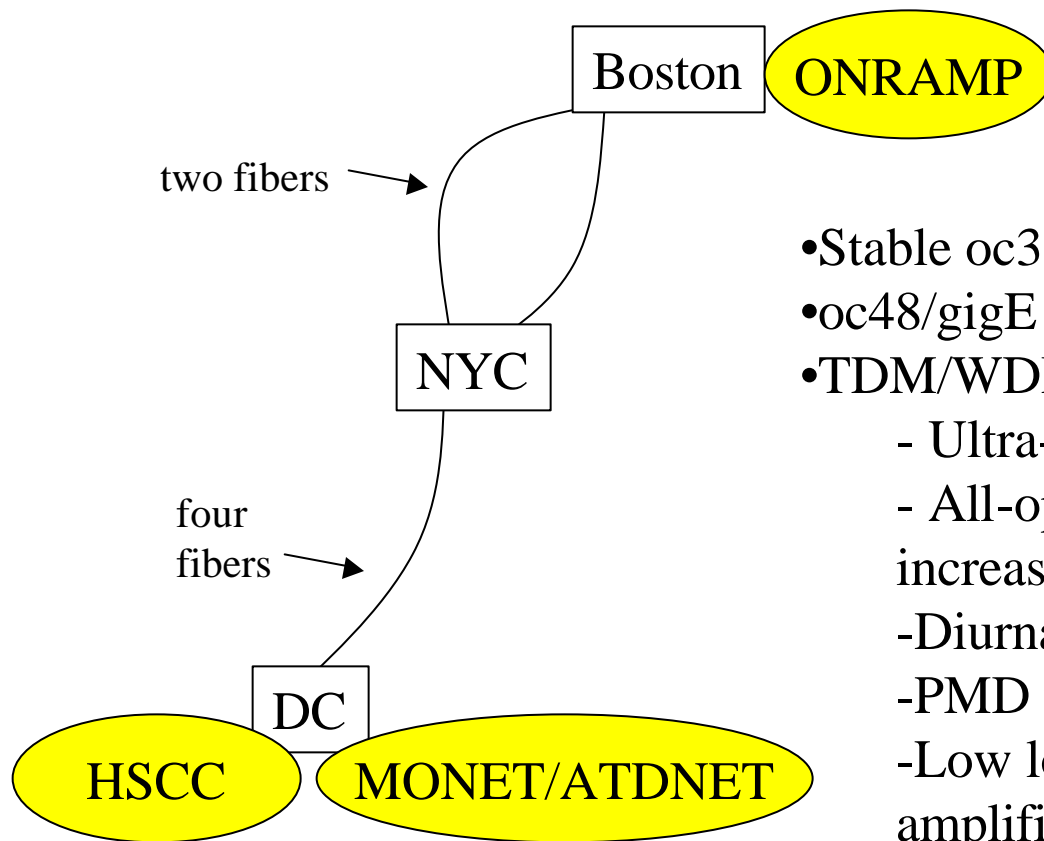
- Stable oc3 channel
- oc48/gigE channel ⟹ for apps
- TDM/WDM Experiments over installed fiber:
  - Ultra-short-pulse 100 Gbps transmission
  - All-optical R^2 regeneration over increasing spans (incl. loopbacks)
  - Diurnal clock recovery investigations
  - PMD tracking/mitigation techniques
  - Low loss window extensions (Raman amplifiers)

# Recent SuperNet Experiments and Demonstrations

- 5x270 Mbps HDTV/POS transmission over 300 km

- 1.2 Gbps TCPIP between desktops; POS 300 km

- 1.5 Gbps HDTV/ATM Transmission over 500 km

- 600 km gbE over MONET/ATDNET

- 10 Gbps dynamic path set up over MONET/ATDNET

- Optical mid-span meet: multi-vendor protection switching demonstration

- Automatic optical layer topology discovery

100's of Gigabit Desktops connected over SuperNet

- Currently targeting 5 campuses (MIT, USC, CMU, UWash,Berkeley)

- Discussions underway with a number of equipment vendors (desktop machines, NIC cards, gE switches, routers)

- Designing a qualification test

- Goal: approximately 50 desktops per each of 5 campuses, 2Q '00

Networking Radars

•Distributed radar control
•Remote data viewing and processing

•New operational paradigm
•Bring down cost/time of research (e.g. design of next generation aircraft one radar)
•Training
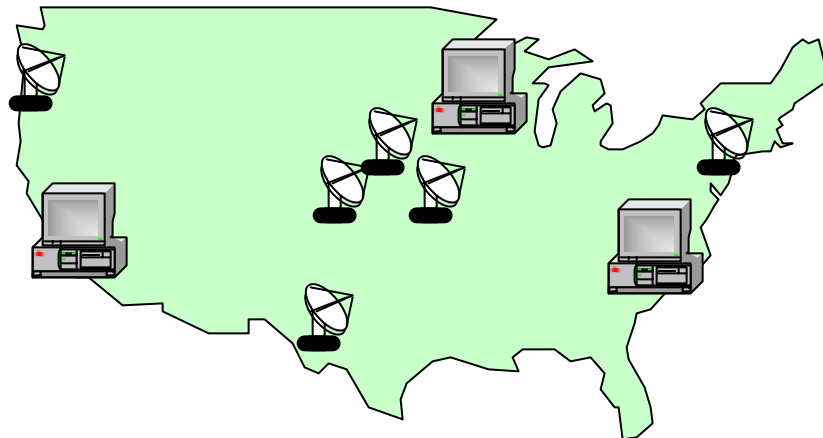
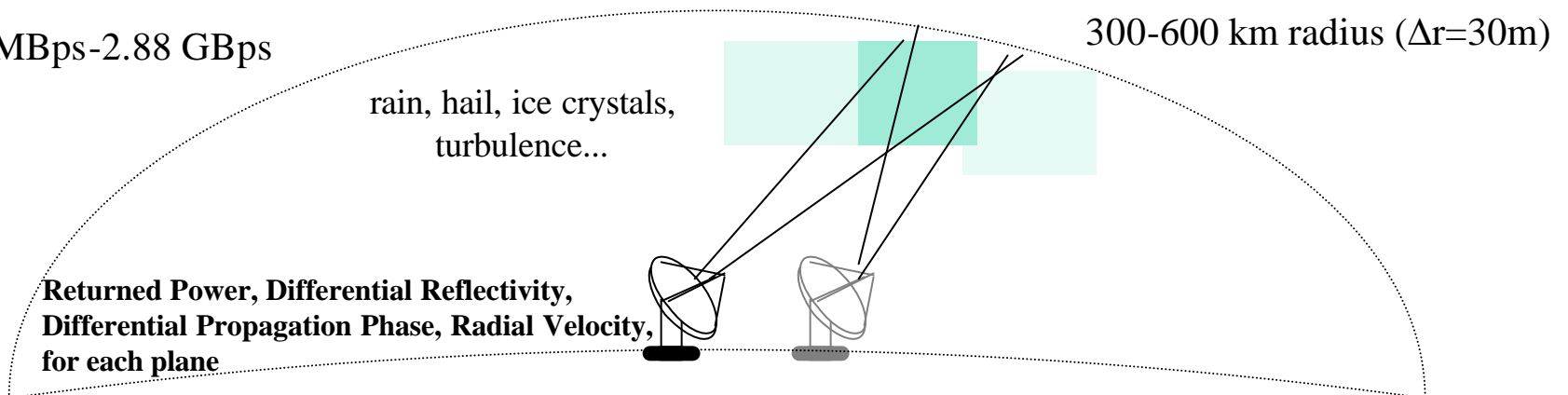# CSU-CHILL Radar for Remote Sensing and Meteorological Analysis

to the network

EM transparent dome,
trailer, radar -- mobile

240 MBps-2.88 GBps

polarimetric,
Doppler radar

300-600 km radius (Δr=30m)

rain, hail, ice crystals,
turbulence...

Returned Power, Differential Reflectivity,
Differential Propagation Phase, Radial Velocity,
for each plane

MATISSE

Networking of sites for testing, characterization, fabrication, users

Example of characterization setup : Computer Microvision Workstation
Characterize MEMS devices by applying cw signal (variable amplitude/freq.)
Optically monitor device response over varying focal planes
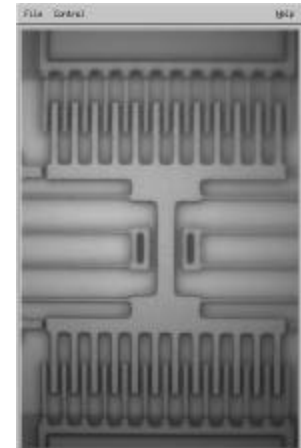
Acoustic/vibrational isolation chamber

Waveform Generator:
• 12-bit waveform generation
• MHz frequencies with mHz resolution
• flexible stroboscopic control

Scientific Microscope:
• ultra-high resolution motor control
• stroboscopic LED illumination

CCD camera system:
• Megapixel camera & frame grabber

typical dataset 10 Gbytes

MIT, CMU, Berkeley, LBL

Video Blanket

... see, remember and understand everything ...

- **Enhanced visualization**

  - Summarize live video from thousands of cameras into a few integrated displays.

  - Video from multiple cameras overlaid in real time over 3D site models to provide scene context.
    - Fly in for best perspective for objects of interest

  - Track events/ people across cameras.

- **Active video surveillance**

  - Close up views co-registered with wide angle views.

  - Virtual walk around of static/ mobile objects of interest

  - Detect/ recognize people (or vehicle) by comparing face/ iris (or license plate) to a database.

- **Analysis of events**

  - Motion tracks, who was where, when and went where.

Video Tracking using a panning camera: Dynamic Mosaic Videos

**Original Video
Tracking Suspect Along Street**

prc_c.avi

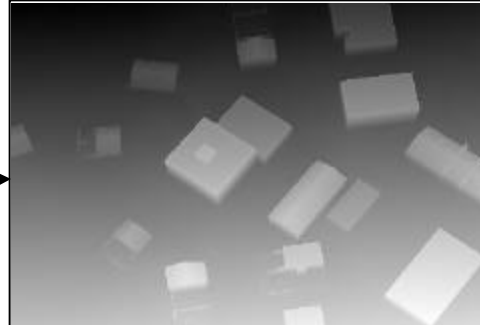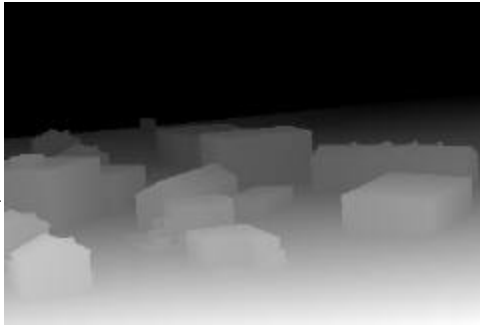**Dynamic Mosaic
Video**

prdyn.avi

**Synopsis
Mosaic**

Visualizing disjoint views

Input data

Re-projection to new view-point after alignment of image frames to 3D model

3D Model

Merging projected frames 1 and frames 2 to create a composite frame from a new view-point

Frame 1

Frame 2

HUBS Telemedicine Application (UPMC, Johns Hopkins)

Beyond text based electronic patient records and proprietary picture archiving/communication systems

- users: radiologists, clinicians, researchers, educators
- *1 image 1 - 10 MegaBytes*
- *one study = 2 to 100 images or 20-100 MB/study*
- *Hopkins and UPMC 700,000 studies/year*
- *35 terabytes/yr or 15 terabytes/yr compressed*

- *3000 new studies/day requires*
- *50 Mbps, x 2 old studies/new studies, 3x clinical transaction = 300 Mbps*

- *Use next-generation VPN service to tie together multiple archiving sites together with users and demonstrate technical feasibility as well as user acceptance*

NGI Multicast Applications and Architecture
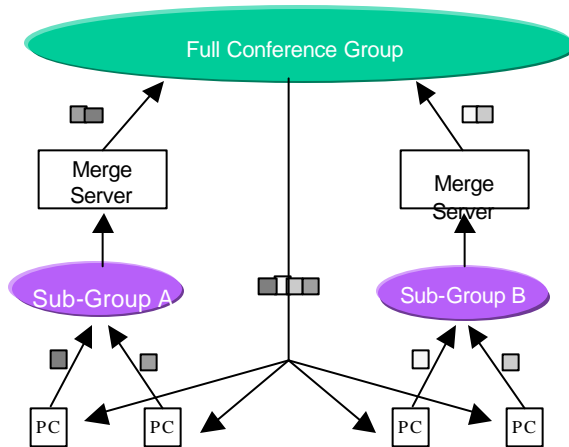
New Ideas

- **Digital Amphitheater** – use dynamic video merge servers in a large-scale multicast architecture to meaningfully conference on the order of <u>a thousand participants</u>, with wide range of participant-controls for the display.

- **Multicasting HDTV** – use RTP and multicast to transmit digital television signals and related data objects on NGI.

Impact

- Push beyond commercial limits of network video.
- Prototype high definition multimedia while retaining economic benefits of commodity computers .
- In collaboration with Corporation for Public Broadcasting, bring about technology transfer between broadcast DTV and broadband networking.

Schedule

USC ⁄ Information Sciences Institute: Allison Mankin
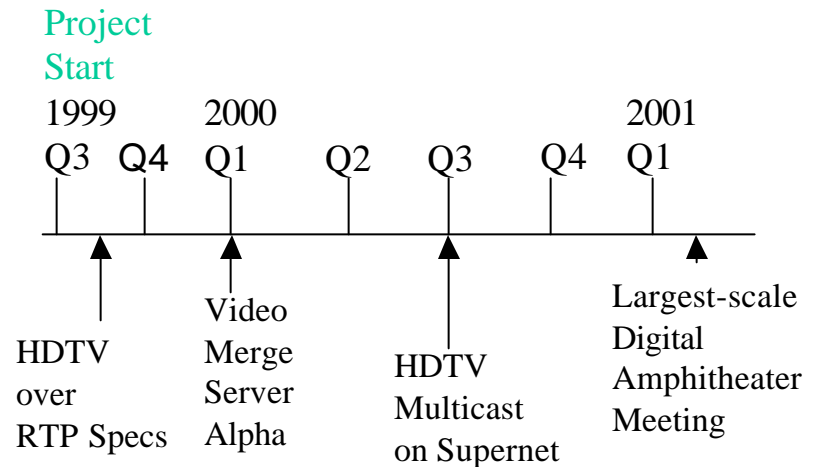
Digital Earth

Open, distributed, scalable multi-resolution 3-D representation of the earth into which massive quantities of geo-referenced information can be embedded.

•Use Domain Name System to develop a hierarchy of servers responsible for geographic cells of earth.
•Enhance today's text-indexing with geographic indexing web to geographically indexed.
• With Virtual Reality Modeling Language (VRML), so with standard browser with plug-in & ~ 50 Mbps, navigate the 3-D model.
•Collaboration between SRI, Planet9 Studios, Sprint.

Infrastructure: .geo domain

- Use DNS to encode latitude/longitude for any element in a hierarchical scheme.
- minutes.degrees.tendegrees.geo
- e.g. 37e47n.1e5n.10e20n.geo